
 M P E R E H

 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

ch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
 Run on: Tue Dec 15 18:30:42 1998; Maspar time 1134.44 Seconds
 Tabular output not generated. 1365.449 Million cell updates/sec

Title: >US-09-109-864-1
 Description: (1-865) from US09109864.seq
 Perfect Score: 865
 N.A. Sequence: 865
 Comp: 1 ttgttgagggcagcagaga.....aaaaaaaaaaaaaaaaaaaaa 865
 aaacacccgcgcgtctct.....tttttttttttttttttt

Scoring table:
 Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 2275026 seqs, 895388244 bases x 2

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: emb1-est55

Database: genbank-est107

1:em_est1 2:em_gss1 3:em_gss2 4:em_gss3
 5:gb_est1 6:gb_est10 7:gb_est11 8:gb_est12 9:gb_est13
 10:gb_est14 11:gb_est15 12:gb_est16 13:gb_est17
 14:gb_est18 15:gb_est19 16:gb_est20 17:gb_est21
 18:gb_est21 19:gb_est3 20:gb_est4 21:gb_est5 22:gb_est6
 23:gb_est7 24:gb_est8 25:gb_est9 26:gb_gss1 27:gb_gss2
 28:gb_gss3 29:gb_gss4

Statistics: Mean 12.059; Variance 7.331; scale 1.645

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	IB	Description	Pred. No.
1	463	53.5	496	24	AA311108	EST181872 Jurkat T-cell	2.06e-287
2	336	38.8	411	16	H30148	yo59h05.rl Homo sapien	2.64e-199
3	319	36.9	422	16	H27468	yo59h05.sl Homo sapien	1.39e-187
4	221	25.5	227	25	AA349100	EST55685 Infant adrena	1.44e-120
5	120	13.9	660	22	AA020729	ze63all.rl Soares reti	2.17e-53
6	80	9.2	434	22	AA001029	ze46b12.rl Soares reti	4.14e-28
7	61	7.1	252	12	AA0754459	97SN1787 Rice Immature	9.89e-17
8	55	6.4	252	12	AA754459	97SN1787 Rice Immature	2.79e-13
9	51	5.9	448	5	R37188	yh96al2.rl Homo sapien	4.94e-11
10	50	5.8	341	12	C84867	Dictyostelium discoide	1.77e-10
11	50	5.8	492	17	AI128239	qc33f09.xl Soares_preg	1.77e-10
12	49	5.7	247	12	AA754458	97SN1784 Rice Immature	6.32e-10
13	49	5.7	558	7	C24325	Dictyostelium discoide	6.32e-10

14	48	5.5	151	14	C90955	Dictyostelium discoide	2.23e-09
15	48	5.5	241	14	C90027	Dictyostelium discoide	2.23e-09
16	48	5.5	252	14	C89779	Dictyostelium discoide	2.23e-09
17	48	5.5	359	14	C91502	Dictyostelium discoide	2.23e-09
18	48	5.5	578	15	C94202	Dictyostelium discoide	2.23e-09
19	48	5.5	785	26	B21585	T28023-Sp6 TAMU Arabid	2.23e-09
20	47	5.4	125	22	N97843	2135C3 czapFD2.1, De	7.83e-09
21	47	5.4	166	11	C84166	Dictyostelium discoide	7.83e-09
22	47	5.4	184	14	C92535	Dictyostelium discoide	7.83e-09
23	47	5.4	241	7	C24389	Dictyostelium discoide	7.83e-09
24	47	5.4	259	23	AA137934	mr01a08.rl Soares mous	7.83e-09
25	47	5.4	267	23	AA231371	mw30907.rl Soares mous	7.83e-09
26	47	5.4	268	14	C91192	Dictyostelium discoide	7.83e-09
27	47	5.4	302	9	AA607600	vo48e06.rl Barstead mo	7.83e-09
28	47	5.4	320	18	AI154582	ud33a01.rl Soares mous	7.83e-09
29	47	5.4	339	7	C23862	Dictyostelium discoide	7.83e-09
30	47	5.4	396	14	AA980543	ua42e08.rl Soares mous	7.83e-09
31	47	5.4	405	18	AA180802	ub76e02.rl Soares mous	7.83e-09
32	47	5.4	476	12	AA747432	nx88f03.sl NCI_CGAP_GC	7.83e-09
33	47	5.4	580	14	C91340	Dictyostelium discoide	7.83e-09
34	47	5.4	614	14	C91555	Dictyostelium discoide	7.83e-09
35	46	5.3	183	14	C93036	Dictyostelium discoide	2.72e-08
36	46	5.3	200	15	C93770	Dictyostelium discoide	2.72e-08
37	46	5.3	224	11	C84161	Dictyostelium discoide	2.72e-08
38	46	5.3	247	12	AA754458	97SN1784 Rice Immature	2.72e-08
39	46	5.3	260	15	AI049844	an21e05.xl Gessler Wil	2.72e-08
40	46	5.3	341	12	AA748451	oa56e01.sl NCI_CGAP_GC	2.72e-08
41	46	5.3	378	18	AI146542	qb45a12.xl NCI_CGAP_Br	2.72e-08
42	46	5.3	437	15	C94484	Dictyostelium discoide	2.72e-08
43	46	5.3	450	14	C89917	Dictyostelium discoide	2.72e-08
44	46	5.3	459	24	AA266894	m298h12.rl Soares mous	2.72e-08
45	46	5.3	821	26	B21581	T25117-Sp6 TAMU Arabid	2.72e-08

ALIGNMENTS

1	RESULT	AA311108	496 bp	mrna	EST	19-APR-1997
	LOCUS	EST181872	Jurkat T-cells	V Homo sapiens	CDNA 5' end, mRNA sequence.	
	DEFINITION	AA311108				
	ACCESSION	AA311108				
	NID	g1963435				
	KEYWORDS	EST.				
	SOURCE	human.				
	ORGANISM	Homo sapiens				
	REFERENCE	1 (bases 1 to 496)				
	AUTHORS	Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Whai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S., Glodek,A., Gnehm,C.D., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palancques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Bednarek,P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Diame,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.D., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.				
	TITLE	Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence				
	JOURNAL	Nature 377 (6547 Suppl), 3-174 (1995)				
	COMMENT	96026280				
	OTHER ESTS	THC124424				
	Contact:	Kerlavage, AR				
	Bioinformatics					
	The Institute for Genomic Research					

9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlavet@tigr.org
 For clone availability, additional sequence and expression
 information related to this EST, please check the TIGR Human Gene
 Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
 Seq primer: M13 Reverse.

FEATURES

Location/Qualifiers
 1..496
 /organism="Homo sapiens"
 /note="vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI"
 /db_xref="ATCC (inhost):157086"
 /db_xref="taxon:9606"
 /clone_lib="Jurkat T-cells v"
 /cell_type="T-lymphocyte"
 <1..5496
 98 a 143 c 151 g 103 t 1 others

BASE COUNT

Query Match 53.5%; Score 463; DB 24; Length 496;
 Best Local Similarity 99.2%; Pred. No. 2.06e-287;
 Matches 482; Conservative 0; Mismatches 1; Indels 3; Gaps 3;
 5 TTGTGGAGGCGACGAGAGTACCCAGCTGGACATCTTCTGCTGAGGCCCGAGG 64
 1 ttgtggagggcagagagagccagctggacatcttcttctgctgagcccgagg 60
 65 CTGAGGTGCGCTCTCTCATCTCTCCAGGCTCTCGCTGGCAGGTCTATCAT 124
 1 ctgaggtgctctctctcatctctccaggtctctgctggcaggtctatcata 120
 125 GTACGGGACTGCTTCTGCAGAGCGGAGGATTTACTGTAGCTGAGGACAGGCT 184
 1 gtacgggactgcttctgcagagcgaggatcttactgtgagctgagggaccagct 180
 185 GCCATCTCTCTGTGACTCAGGCGCTCTTCGAGACAACTCTGGCTGCTTCC 244
 1 gccatctctctgtgactcaggcgcttcttcgagacaaactctggctgcttcc 240
 181 gccatctctctgtgactcaggcgcttcttcgagacaaactctggctgcttcc 240
 245 CGCTGGGGCAGAGAACTGATCTACGCCCTCTCTCTTTACCCCGAGAGATCTTT 304
 1 cgctggggcagagaaactgactcagccctctctctttaccccgagagatcttt 300
 305 GAGGTGCTCTCTGTCCAGAGGAGGCTGAGCTGCGCTCAATGGCAGGGGCTG 364
 1 gagggtgctctctgttccagagggagggctgagctgcgctcaatggcaggggctg 360
 365 GGGGCCACGACATGAACACAGCCCTGAGAGCTGCGGGAGCTCCGGATCAGTGA 424
 1 gggggccacgacatgaacacagccctgagagctgcgggagctccggatcagtga 420
 425 AGTGTCAGCTCTACTGTGTCCACTCTCTGA-GGATGTTCCAGGGAATACCGCAGAA 483
 1 agtgtcagctctactgtgtccactctctga-ggatttccagggaggttccagg-aaatacgc-cagaaa 478
 484 ACAGA 489
 1 acaaga 484

RESULT 2
 LOCUS H30148 411 bp mRNA EST 16-AUG-1995
 DEFINITION yo59h05.r1 Homo sapiens cDNA clone 182265 5'
 ACCESSION H30148
 NID 9901058
 EST.
 KEYWORDS human clone-182265 library=Soares breast 3NdbEst vector=pt7t3D
 (Pharmacia) with a modified polylinker host=DH10B (ampicillin
 resistant) primer=M13R1 Rsitel-Not I Rsite2-Eco RI Adrit human.
 1st strand cDNA was primed with a Not I - oligo(GT) pr.mer [5',
 TGTTACCATCTGAGTGAGGAGCGGCCCGCCCTTTTITTTTTTTT 3'],

double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia),
 digested with Not I and cloned into the Not I and Eco RI sites of a
 modified pTT73 vector (Pharmacia). Library was then through one round
 of normalization to a Cot = 20. Library constructed by Bento Soares
 and M.Fatima Bonaldo.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Eumetazoa; Bilateria; Coelomata;
 Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
 Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;
 Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 411)
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevas, E., Waterston, R., Williamson, A., Wohlmann, P. and
 Wilson, R.
 TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT

Contact: Wilson RK
 WashU-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 High quality sequence stops: 264
 Source: IMAGE Consortium, LLNL
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (image.llnl.gov) for further information.
 FEATURES source
 1..411
 /organism="Homo sapiens"
 /clone="182265"
 90 a 113 c 121 g 79 t 8 others

Query Match 38.8%; Score 336; DB 16; Length 411;
 Best Local Similarity 95.7%; Pred. No. 2.64e-199;
 Matches 360; Conservative 0; Mismatches 13; Indels 3; Gaps 3;
 Db 1 CAGGGCTCTNACAGGCGAGTCTATCATAGTACGGGANTGGCTTNCAGAGCGGAG 60
 1 cagggtctctnaccaggcgagtctatcatagtagcgggagntggcttncagagcgag 150
 61 CATTTTACTGTGAGCTGAGGACCGAGCTNCCCATGCTCTGTGACACTCAGGCGCTCC 120
 1 catTTTactgtgagctgaggacccgagctncccatgctctgtgacactcaggcgctcc 210
 151 cattttactgtgagcctgagggaccaggtgcccatgtctctgtgacactcaggcgctcc 210
 121 TTGCGAGACAGAACTCTGCGCTGGATCTCCGCTGGGGCAGAGAACTGATCTCAGCC 180
 1 ttgCGAGACAGAACTCTGCGCTGGATCTCCGCTGGGGCAGAGAACTGATCTCAGCC 270
 211 ttgcgagacagaactctggcctggatctcccgctggggcagagaaagtatcagcc 240
 181 CCCTTCTCTTTTACCNCAGAGATCTTTGAGTGTGCTGCTCTCTTCCAGAGGAGGG 240
 1 cccttctctcttttaccncagagatctttgagtggtgtgctgctcttccagagagg 330
 271 ccttctctcttttaccncagagatctttgagtggtgtgctgctcttccagagagg 300
 241 CTGAGAGCTGGGCTCAATGGGCGAGGCGCTGGGGCCACCAAGCAATGAACAGGAGGCC 388
 1 ctgagagctgggctcaatgggCGAGGCGCTGGGGCCACCAAGCAATGAACAGGAGGCC 360
 331 ctgaagctgggctcaatgggCGAGGCGCTGGGGCCACCAAGCAATGAACAGGAGGCC 388
 301 TGGAGCAGCTGGGNGACTCCGGATCAGTGGAGTATTCAGCTCTACTGTGCTCCACTCT 360
 1 tggagcagctgggngactccggatcagtggaagtgtcagctctactgtgtccactct 448
 389 tggagcagctgggngactccggatcagtggaagtgtcagctctactgtgtccactct 448
 361 NA-GGATGGTCCAGG 375
 1 na-ggagtggttccagg 464
 449 gaagtggttccagg 464

RESULT 3 H27468 422 bp mRNA EST 16-AUG-1995
 LOCUS yo59h05.s1 Homo sapiens cDNA clone 182265 3',
 DEFINITION

W P E R L A H (TM)

Release 3.1A John F. Collins, Blocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Dec 15 15:55:11 1998; MasPar time 5.67 Seconds
629.394 Million cell updates/sec

Tabular output not generated.

Title: >US-09-109-864-2
Description: (1-133) from US09109864.pcp
Perfect Score: 980
Sequence: 1 MSPRLVPCSHALPQGLSPG.....LEQLRELISGSVQLTCVHS 133

Scoring table: PAM 150
Gap 11

Searched: 74019 seqs, 26840295 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot35
1:swissprot

Statistics: Mean 44.065; Variance 78.778; scale 0.559

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	174	17.8	144	1	LEG5_RAT	2.71e-14
2	146	14.9	354	1	LEG9_RAT	2.88e-09
3	140	14.3	316	1	LEG8_HUMAN	3.11e-08
4	140	14.3	353	1	LEG8_MOUSE	3.11e-08
5	136	13.9	135	1	LEG7_HUMAN	1.49e-07
6	125	12.8	135	1	LEG7_HUMAN	9.88e-06
7	112	11.4	279	1	LEG3_CAEEL	1.12e-03
8	112	11.4	289	1	LEG3_CAEEL	1.12e-03
9	109	11.1	316	1	LEG8_RAT	3.21e-03
10	109	11.1	324	1	LEG4_RAT	3.21e-03
11	106	10.8	448	1	GPT_YEAST	9.03e-03
12	99	10.1	134	1	LEG4_CHICK	9.41e-02
13	98	10.0	308	1	NSP_PSEST	1.30e-01
14	97	9.9	184	1	YRNA_CAEEL	1.80e-01
15	95	9.7	746	1	PEEA_PSEAE	3.42e-01
16	94	9.6	135	1	LEG7_RAT	4.70e-01
17	93	9.5	493	1	GSPE_ECOLI	6.43e-01
18	93	9.5	1522	1	CIN1_LOLAL	6.43e-01
19	92	9.4	437	1	SECY_STRGR	8.79e-01
20	92	9.4	583	1	CFAL_HUMAN	8.79e-01
21	92	9.4	620	1	KFCF_ECOLI	8.79e-01
22	90	9.2	510	1	EMRB_HAEIN	1.63e+00
23	90	9.2	761	1	PQOF_KLEPN	1.63e+00

24	89	9.1	225	1	NTCA_SYNY3	GLOBAL NITROGEN REGULA	2.21e+00
25	89	9.1	323	1	LEG4_HUMAN	GALECTIN-4 (LACTROSE-BI	2.21e+00
26	89	9.1	563	1	YEOK_SCHPO	HYPOTHETICAL PROTEIN C	2.21e+00
27	89	9.1	976	1	FIBP_ADEB3	FIBER PROTEIN.	2.21e+00
28	88	9.0	137	1	YSCB_YEREN	PUTATIVE YOP PROTEINS	2.98e+00
29	88	9.0	218	1	SIR_DESVH	SULFITE REDUCTASE, ASS	2.98e+00
30	87	8.9	198	1	COAT_BSMV	COAT PROTEIN (CAPSID P	4.02e+00
31	87	8.9	282	1	SPRE_RAT	SEPIA FERIN REDUCTASE	4.02e+00
32	87	8.9	343	1	HOLA_ECOLI	DNA POLYMERASE III, DE	4.02e+00
33	87	8.9	786	1	YBIO_ECOLI	HYPOTHETICAL 86.8 KD P	4.02e+00
34	87	8.9	1513	1	RPOD_ORYSA	DNA-DIRECTED RNA POLYM	4.02e+00
35	87	8.9	1527	1	RPOD_MAIZE	DNA-DIRECTED RNA POLYM	4.02e+00
36	87	8.9	1742	1	GUNA_CALSA	ENDOGLUCANASE A PRECUR	4.02e+00
37	86	8.8	481	1	ATZB_PESD	HYDROXYLASE HYDROL	5.40e+00
38	85	8.7	223	1	NTCA_ANASP	GLOBAL NITROGEN REGULA	7.24e+00
39	85	8.7	244	1	LEG3_CRILLO	GALECTIN-3 (GALACTOSE-	7.24e+00
40	85	8.7	424	1	MTBL_BACSH	MODIFICATION METHYLASE	7.24e+00
41	85	8.7	495	1	ACHE_RAT	NEURONAL ACETYLCHOLINE	7.24e+00
42	85	8.7	725	1	GUNG_CLOCE	ENDOGLUCANASE G PRECUR	7.24e+00
43	85	8.7	868	1	MCE_ASFB7	MRNA CAPPING ENZYME (M	7.24e+00
44	85	8.7	879	1	GUNI_CLOTM	ENDOGLUCANASE I PRECUR	7.24e+00
45	85	8.7	1108	1	MAN2_DROME	ALPHA-MANNOSIDASE II (7.24e+00

ALIGNMENTS

RESULT 1
ID LEG5_RAT STANDARD; PRT; 144 AA.
AC P47967:
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE GALECTIN-5 (RL-18).
GN LGALS5.
OS RATIUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-RETICULOCYTES;
RX MEDLINE; 95197487.
RA GITT M.A., WISER M.F., LEFFLER H., HERRMANN J., XIA Y.-R.,
RA KASSA S.H., COOPER D.N.W., JUSIS A.J., BARONDES S.H.,
RL J. BIOL. CHEM. 270:5032-5038(1995).
CC - FUNCTION: MAY FUNCTION IN ERYTHROCYTE DIFFERENTIATION.
CC - SUBUNIT: MONOMER.
CC - TISSUE SPECIFICITY: ERYTHROCYTES.
CC - SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
DR EMBL; L36862; G727176.
DR EMBL; L21711; G785053.
KW GALAPTIN; LECTIN; ACETYLATION.
FT INIT_MET 0 0
FT MOD_RES 0 0
FT BINDING 76 82
FT CONFLICT 127 135
FT SEQUENCE 144 AA: 16065 MW; 5C7CC533 CRC32:
Query Match 17.8% Score 174; DB 1: Length 144;
Best Local Similarity 28.0% Pred. No. 2.71e-14;
Matches 37: Conservative 38; Mismatches 49; Indels 8; Gaps 6;

Db	10	PNLAVFFTSIPNGLVPSKIVTSIGVVLSDAKRFQINLRC-GGDIAFHLNPRFENAVVR	68
QY	3	PRLEVPCHALPQGLSPGVIIIRGLVLOEPKHFTVSLRQAAHAPVTLRAADRTLAW	62
Db	69	NTQINNSWGPESRLPGSMFSGQRF-SWILCEGRCFKVAVDGGHICYSRLNLPD	127
QY	63	---ISR-WG-QKKLISAPFLFYP-QRFFVLLLFQEGGLKALNGGIGATSNQOALEQ	116
Db	128	INTLEVAGDTQL	139
QY	117	IRELRISGSVQL	128

RESULT	3
ID	LEG8_HUMAN STANDARD; PRT; 316 AA.
AC	O00214;
DT	01-NOV-1997 (REL. 35, CREATED)
DI	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT	15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE	GALACTIN-8 (PROSTATE CARCINOMA TUMOR ANTIGEN) (PCTA-1).
DN	LGALS8.
GN	HOMO SAPIENS (HUMAN).
OS	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC	EUTHERIA; PRIMATES.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=PROSTATE;
RX	MEDLINE; 96293510.
LA	SU Z LIN J., SHEN R., FISHER P.E., GOLDSTEIN N.I., FISHER P.B.;
PL	PROC. NATL. ACAD. SCI. U.S.A. 93:7252-7257(1996).
CC -1-	SUBCELLULAR LOCATION: CITOPASMIC (PROBABLY).
CC -1-	TISSUE SPECIFICITY: SELECTIVE EXPRESSION BY PROSTATE CARCINOMAS
CC	VERSUS NORMAL PROSTATE AND BENIGN PROSTATIC HYPERTROPHY.
CC	-1- DOMAIN: CONTAINS TWO HOMOLOGS BUT DISTINCT CARBOHYDRATE-BINDING
CC	DOMAINS.
CC -1-	SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
DR	EMBL; L78132; G1932712; .
DR	PROSITE; PS00309; LECTIN GALACTOSIDE; I.
KW	GALAPTIN; LECTIN; REPEAT; ANTIGEN.
FT	DOMAIN 1 153 GALAPTIN 1.
FT	DOMAIN 154 184 LINKER.
FT	DOMAIN 185 315 GALAPTIN 2.
FT	BINDING 248 234 BETA-GALACTOSIDE (BY SIMILARITY).
SQ	SEQUENCE 316 AA; 35588 MW; EB4D9A3D CRC32;
Query Match 14.3%; Score 140; DB 1; Length 316;	
Best Local Similarity 24.8%; Pred.No. 3.11e-08;	
Matches 34; Conservative 4; Mismatches 56; Indels 7; Gaps	
Dd	179 TPQLSLFFAARLTNPMPGRTVVVOGEVNANAKSENVLLACKSKNDIALHLNPRLNKAF 238 ::: : :::: : :::: : :::: : :::: : :::: : :::: : :::: : :::: :
Qy	2 SPRLVEPCSHALPGGLSPGVIIIRGLVLQEPKHETSU-RDQAHAHPVTLRASFADRTL 60 :
Dd	239 VRNSFLOESNGEERMITSFPSPGMFYFMYYVDREKVAVNGVHSLEYKKHFKE-LS 297 :
Qy	61 AWIS--R-WGRKLKSAPLEPYQRFYFEVLIVFOEGGLKALANG-QGLGATSMNQOALE 115 :
Dd	298 SIDTLEINGDIMLEVRS 315 :
Qy	116 QLRELRISSGVQLYCVMYS 133 :
RESULT	4
ID	LEG9_MOUSE STANDARD; PRT; 353 AA.
AC	O08573; O08572;
DT	01-NOV-1997 (REL. 35, CREATED)
DI	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE	15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE	GALACTIN-9.
GN	LGALS9.
OS	MUS MUSCULUS (MOUSE).
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC	EUTHERIA; RODENTIA.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CD-1; TISSUE=SMALL INTESTINE, AND KIDNEY;
RX	MEDLINE; 97190351.
RA	WADA J., KANWAR Y.S.;
JL	J. BIOL. CHEM. 272:6078-6086(1997).
RN	[2]
RP	CHARACTERIZATION.
RX	MEDLINE; 97298141.
RA	WADA J., OTA K., KUMAR A., WALLNER E.I., KANWAR Y.S.;
JL	J. CLIN. INVEST. 99:2452-2461(1997).
CC -1-	FUNCTION: BINDS GALACTOSIDES. MAY PLAY A ROLE IN THYMOCYTE- EPITHELIAL INTERACTIONS RELEVANT TO THE BIOLOGY OF THE THYMUS.

```

CC  -1- TISSUE SPECIFICITY: MAINLY IN STRATIFIED SQUAMOUS EPITHELIUM.
CC  -1- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
DR  EMBL; L07769; G182132; -.
DR  EMBL; U06643; G458703; -.
DR  AARRUS/GHENT-2DPAGE; 17; IEF.
DR  MIM; G00615; -.
DR  DR PROSITE; PS00309; LECTIN_GALACTOSIDE; 1.
KW  GALAPTIN; LECTIN.
FT  INIT.MET 0
FT  BINDING 69 75 BETA-GALACTOSIDE (POTENTIAL).
SQ  SEQUENCE 135 AA; 14944 MW; 33656F93 CRC32;

Query Match 13.9%; Score 136; DB 1; Length 135;
Best Local Similarity 28.1%; Pred. No. 1.49e-07;
Matches 38; Conservative 39; Mismatches 47; Indels 11; Gaps

Db 2 NYPKSLSPGIRGTVLRIGLVPPNASEFHNLLCGEQGSDAALHFNPRLDTSEVF 61
   :|| :||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 6 EYVCSHALPGSLSPGVIIYRGLVQLPQKHTVSL---RQAAHAPVTLRASFAADRTLAW 62
   :|| :|| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 62 NSKEQGSWGREERGPGVPFOR-GQPF-EVILLASDDGFEK-AVYVGDACYNHFNRLPLARY 118
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 63 ISR----WG-QKKLIAPFLEYPQRFEEVLLFQEGGLKALNGQGLGATSMQQALEDQL 117
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 119 RLVEVGQDVQLDSVR 133
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Qy 118 RELRIGSVOLYCVH 132
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

```

RESULT	6	LEG9_HUMAN	STANDARD;	PRT;	355 AA.
ID	AC	000182;	O14532;		
DT	DT	11-NOV-1997	(REL. 35, CREATED)		
DT	DT	15-JUL-1998	(REL. 36, LAST SEQUENCE UPDATE)		
DT	DT	15-JUL-1998	(REL. 36, LAST ANNOTATION UPDATE)		
DE	DE	GALECTIN-9 (HOM-HD-21).			
GN	GN	LGALS9.			
OS	OS	HOMO SAPIENS (HUMAN).			
OC	OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TE			
OC	OC	EUTHERIA; PRIMATES.			
		[1]			
RN	RN	SEQUENCE FROM N.A. (SHORT FORM).			
RP	RP	TISSUE=SPLEN;			
RC	RC				

[2]
SEQUENCE FROM N.A. (LONG FORM).
TISSUE-GASTRIC CARCINOMA;
KATO S.;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -I- FUNCTION: BINDS GALACTOSIDES, MAY PLAY A ROLE IN THYMOCYTE-
CC EPIHELIAL INTERACTIONS RELEVANT TO THE BIOLOGY OF THE THYMUS.
CC -I- TISSUE SPECIFICITY: PERIPHERAL BLOOD LEUKOCYTES AND LYMPHATIC
CC TISSUES. OVEREXPRESSED IN HODGKIN'S DISEASE TISSUE.
CC -I- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
CC DOMAINS.
CC -I- ALTERNATIVE PRODUCTS: AT LEAST TWO FORMS ARE PRODUCED BY
CC ALTERNATIVE SPLICING.
CC -I- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
DR EMBL: Z49107; E310691; -
DR EMBL: AB006782; D1023026; -
DR MTM: 601879; -
DR PROSITE: PS00309; LECTIN_GALACTOSIDE; 2.
KW GALAPTIN; LECTIN; REPEAT; ALTERNATIVE SPLICING.
FT DOMAIN 1 148 GALAPTIN 1.
FT DOMAIN 149 206 LINKER.
FT DOMAIN 207 355 GALAPTIN 2.
FT BINDING 82 88 BETA-GALACTOSIDE (BY SIMILARITY).
FT BINDING 287 293 BETA-GALACTOSIDE (BY SIMILARITY).
FT VARSPIC 149 180 MISSING (IN SHORT ISOFORM).
FT CONFLICT 88 88 K -> R (IN REF. 1).
FT CONFLICT 135 135 S -> P (IN REF. 1).

```

01-OCT-1996 (REL. 34, CREATED)
01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
PROBABLY 33 KD BETA-GALACTOSIDE-BINDING LECTIN (33 KD GBP).
2K892.1
CAENORHABDITIS ELEGANS.
EUKARYOTA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
[1]
SEQUENCE FROM N.A.
STRAIN=BRISTOL N2;
LLOYD C.;
SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
-!- FUNCTION: BINDS GALACTOSE (BY SIMILARITY).
-!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
EMBL; Z48638; G732606; -.
WORMPEP; ZK892.1; CE01724.
PROSITE; PS00309; LECTIN_GALACTOSIDE; 1.
HYPOTHEETICAL PROTEIN; GALAPTIN; LECTIN; REPEAT.
FT DOMAIN 1 141 GALAPTIN 1.
FT DOMAIN 142 289 GALAPTIN 2.
FT SEQUENCE 289 AA; 32991 MW; 3329858D CRC32;

Query Hatch 11.4%; Score 112; DB 1; Length 289;
Best Local Similarity 29.3%; Pred. Ho. 1.12e-03;
Matches 12; Conservative 17; Mismatches 12; Indels 0; Gaps 0;

Db 11 IHPYRSKTEREPGQTILRGKTIIDSKRFNINLHKDSP 51
:::| | | | | | | | | | | | | | | | | | | |
5 LEVPCSHALPGQISPGQVIVRGVLQEPKFTVSLRDQAA 45

RESULT 9
ID LEG8_RAT STANDARD; PRT; 316 AA.
AC O62665;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE GALECTIN-8 (30 KD S-TYPE LECTIN) (RL-30).
GN LGALS8.
OS RATUS HORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
[1]
SEQUENCE FROM N.A.
STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;
MEDLINE; 95155445.
RA HADARI Y.R.; FAZ K.; DEREL R.; MESTROVIC T.; ACCILI D.; ZICK Y.;
RL J. BIOL. CHEM. 270:3447-3453(1995).
CC -!- FUNCTION: POSSESS SUGAR BINDING AND HEMAGGLUTINATION ACTIVITY.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
CC DOMAINS.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN LIVER, KIDNEY, CARDIAC MUSCLE,
CC LONG, AND BRAIN.
CC -!- DEVELOPMENTAL STAGE: VERY LOW LEVELS IN WHOLE EMBRYOS, HIGH LEVELS
CC IN ADULT TISSUES.
CC -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
DR EMBL; U09824; G717032; -.
DR PROSITE; PS00309; LECTIN_GALACTOSIDE; 1.
KW GALAPTIN; LECTIN; REPEAT.
FT DOMAIN 1 153 GALAPTIN 1.
FT DOMAIN 154 184 LINKER.
FT DOMAIN 185 316 GALAPTIN 2.
FT BINDING 248 254 BETA-GALACTOSIDE (BY SIMILARITY).
FT SEQUENCE 316 AA; 36038 MW; 96341BC7 CRC32;

Query Hatch 11.1%; Score 109; DB 1; Length 316;
Best Local Similarity 22.1%; Pred. Ho. 3.21e-03;
Matches 30; Conservative 41; Mismatches 58; Indels 7; Gaps 5;

Db 181 HLSLPTEARLNAMSGPRIVVVKGEYNTNATSNVDLVAGRSRDIALHLPRLNVAFVR 240
QY 4 RLVPQSHALPGQISPGQVIVRGVLQEPKFTVSL-RDQAAHPVTLNRSADFRTLAW 62

```

Db 241 NSFTQDANGBERNITCFPSSGMIFEMILYCDVREKVAVGVHSHLEYKHKFKD--LS51 299
 | : || : | : : || : : : || : | : | : | : | : | : | : | : | : | : | :
 Qy 63 IS--R-WGKKLISAPFETPORFEVLLFQGGGLKALNG-QGIGATSMNOQALEQL 117
 | : || : | : : || : : : || : | : | : | : | : | : | : | : | : | : | :
 Db 300 DTLAVDGDIRLLDVR5 315
 | : || : | : : || : | : | : | : | : | : | : | : | : | : | : | :
 Qy 118 RELRISGVOLYCVH5 133
 | : || : | : : || : | : | : | : | : | : | : | : | : | : | : | :
 RESULT 10
 ID LEG4.RAT STANDARD; PRT; 324 AA.
 AC P38552;
 DT 01-OCT-1994 (REL. 30, CREATED)
 DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE GALECTIN-4 (LACTOSE-BINDING LECTIN 4) (L-36 LACTOSE BINDING PROTEIN)
 GN (L36LBP).
 GE LGALS4.
 OS RATTUS NORVEGICUS (RAT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; RODENTIA.
 RP [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE-INTESTINE;
 RX MEDLINE; 93194902.
 RA BARONDES S.H., LEFFLER H.;
 RA ODA Y., HERRMANN J., GITT M., TURCK C.W., BURLINGAME A.L.,
 RA J. BIOL. CHEM. 268:5929-5939(1993).
 CC -1- FUNCTION: GALECTIN THAT BINDS LACTOSE AND A RELATED RANGE OF
 CC SUGARS.
 CC -1- SUBUNIT: MONOMER.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN FULL-LENGTH FORM IN SMALL
 CC AND LARGE INTESTINE AND STOMACH BUT WAS NOT DETECTED IN OTHER
 CC TISSUES INCLUDING LUNG, LIVER, KIDNEY, AND SPLEN.
 CC -1- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
 CC DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
 DR EMBL: M73553; G294572; -.
 DR PIR; A46631; A46631.
 DR PROSITE; PS00309; LECTIN_GALACTOSIDE; 2.
 KW GALAPTIN; LECTIN; REPEAT.
 FT DOHA1N 1 152 GALAPTIN 1.
 FT DOHA1N 153 177 LINKER.
 FT DOMAIN 178 324 GALAPTIN 2.
 FT BINDING 257 263 BETA-GALACTOSIDE (BY SIMILARITY).
 SQ SEQUENCE 324 AA; 36347 MW; 98702405 CRC32;
 Query Match 11.1%; Score 109; DB 1; Length 324;
 Best Local Similarity 23.7%; Pred.No. 3.21e-03;
 Matches 31; Conservative 38; Mismatches 56; Indels 6; Gaps
 Db 194 VPYVGTGGTARRIIIKGYVLPAKLLIIFKVGSTGDIAPHNPRIGDCVVENSYM 253
 | | : | | : : | | : | | : | | : | | : | | : | | : | | : | : | :
 Qy 7 VPCHAUQGLSGQVLIIVGLVLPKPHFTVSLR-DQAHAHPVILRASFADRTL--AWI 63
 | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | :
 Db 254 NGWSGE-RKIYPNPGACOFFDLSCRTGDFKVFANGHLPDFSHRFQAFQVDMLE 312
 : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | :
 Qy 64 S-RWGKKLISAPF-LFPYQPFVLLFQGGGLKALNGIGATSMNOQALEQLREL 121
 | : || : | : : || : | : | : | : | : | : | : | : | : | : | : | :
 Db 313 IKGDITLSYVO 323
 | : || : | : : || : | : | : | : | : | : | : | : | : | : | : | :
 Qy 122 ISGSVOLYCVH 132
 | : || : | : : || : | : | : | : | : | : | : | : | : | : | : | :
 RESULT 11
 ID GPT-YEAST STANDARD; PRT; 448 AA.
 AC P07286;
 DT 01-APR-1988 (REL. 07, CREATED)
 DT 01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE UDP-N-ACETYLGLUCOSAMINE--DOLICHL-PHOSPHATE N-
 DE ACETYLGLUCOSAMINEPHOSPHATRANSFERASE (EC 2.7.8.15) (GPT) (G1PT)

[illegible]

```

RESULT 15
ID  PFEA_PSEAE      STANDARD;      PRT;      746 AA.
AC  Q05098;
DT  01-JUN-1994 (REL. 29, CREATED)
DT  01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT  01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE  FERRIC ENTEROBACTIN RECEPTOR PRECURSOR.
GN  PFEA.
OS  PSEUDOMONAS AERUGINOSA.
OC  PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
OC  PSEUDOMONADACEAE.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=K407;
RX  MEDLINE: 93I23I48.
RA  DEAN C.R., POOLE K.
RL  J. BACTERIOL. 175:317-324(1993).
CC  -!- FUNCTION: SPECIFIC RECEPTOR FOR THE SIDEROPHORE FERRIC
CC  ENTEROBACTIN.
CC  -!- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC  -!- INDUCTION: REGULATED BOTH BY IRON AND ENTEROBACTIN.
CC  -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
DR  EMBL: M98033; GI51435; -.
DR  PIR: A40636; A40636.
DR  PROSITE: PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR  PROSITE; PS01156; TONB_DEPENDENT_REC_2; I.
KW  OUTER MEMBRANE; IRON_TRANSPORT; TONB_BOX; SIGNAL; RECEPTOR.
FT  SIGNAL 1 25
FT  CHAIN 26 746
FT  FERRIC ENTEROBACTIN RECEPTOR.

```


FT SIMILAR 39 44 TONB BOX.
FT SIMILAR 729 746 TONB C-TERMINAL BOX.
SQ SEQUENCE 746 AA; 80967 MW; 3BC231B9 CRC32;

Query Match 9.7%; Score 95; DB 1; Length 746;
Best Local Similarity 37.8%; Pred. No. 3.42e-01;
Matches 14; Conservative 10; Mismatches 12; Indels 1; Gaps 1;

DB 363 YTATLRDLTAHGEVNLPHLGYEQTLILGSEWTEOKL 399
QY 36 FTVSLRDQAHAHPVTLRASFA-DETLAWISRWGOKKL 71

Search completed: Tue Dec 15 15:55:27 1998
Job time : 16 secs.

W P E R L H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 15 18:30:42 1998; MasPar time 1134.44 Seconds
Tabular output not generated. 1365.449 Million cell updates/sec

Title: >US-09-109-864-1
Description: (1-865) from US09109864.seq
Perfect Score: 865
N.A. Sequence: 1 ttgttgaggcagcagaga.....aaaaaaaaaaaaaaaa 865
Comp: aaacactccctgctctc.....tttttttttttttttttt

Scoring Table: TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 2275026 seqs, 895388244 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: emb1-est55

Database: genbank-est107

5:gb_est1 6:gb_est10 7:gb_est11 8:gb_est12 9:gb_est13
10:gb_est14 11:gb_est15 12:gb_est16 13:gb_est17
14:gb_est18 15:gb_est19 16:gb_est20
18:gb_est21 19:gb_est23 20:gb_est24 21:gb_est25 22:gb_est26
23:gb_est27 24:gb_est28 25:gb_est29 26:gb_gssl 27:gb_gssl2
28:gb_gssl3 29:gb_gssl4

Statistics: Mean 12.059; Variance 7.331; scale 1.645

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	463	53.5	496	24	AA311108 EST181872 Jurkat T-cell	2.06e-287
2	336	38.8	411	16	H30148 yo59h05.r1 Homo sapien	2.64e-199
3	319	36.9	422	16	H27468 yo59h05.s1 Homo sapien	1.39e-187
4	221	25.5	227	25	AA349100 EST55685 Infant adrena	1.44e-120
5	120	13.9	660	22	AA020729 ze63a11.r1 Soares reti	2.17e-53
6	80	9.2	434	22	AA001029 ze63b12.r1 Soares reti	4.14e-28
7	61	7.1	252	12	AA754459 97SN1787 Rice Immature	9.89e-17
8	55	6.4	252	12	AA754459 97SN1787 Rice Immature	2.79e-13
9	51	5.9	448	5	R37188 yb6a12.r1 Homo sapien	4.94e-11
10	50	5.8	341	12	C84867 Dictyostelium discoide	1.77e-10
11	50	5.8	492	17	AA128239 qc33f09.x1 Soares_preg	1.77e-10
12	49	5.7	247	12	AA754458 97SN1784 Rice Immature	6.32e-10
13	49	5.7	558	7	C24325 Dictyostelium discoide	6.32e-10

14	48	5.5	151	14	C90955 Dictyostelium discoide	2.23e-09
15	48	5.5	241	14	C90027 Dictyostelium discoide	2.23e-09
16	48	5.5	252	14	C89779 Dictyostelium discoide	2.23e-09
17	48	5.5	359	14	C91502 Dictyostelium discoide	2.23e-09
18	48	5.5	578	15	C94202 Dictyostelium discoide	2.23e-09
19	48	5.5	785	26	B21585 T28D23-Sp6 TAMU Arabid	2.23e-09
20	47	5.4	125	22	N97843 2135C3 czapFD42.1. De	7.83e-09
21	47	5.4	166	11	C84166 Dictyostelium discoide	7.83e-09
22	47	5.4	184	14	C92535 Dictyostelium discoide	7.83e-09
23	47	5.4	241	7	C24389 Dictyostelium discoide	7.83e-09
24	47	5.4	259	23	AA137934 mw01a08.r1 Soares mous	7.83e-09
25	47	5.4	267	23	AA231371 mw03a07.r1 Soares mous	7.83e-09
26	47	5.4	268	14	C91192 Dictyostelium discoide	7.83e-09
27	47	5.4	302	9	AA607600 vo46e06.r1 Barstead mo	7.83e-09
28	47	5.4	320	18	AI154582 Dictyostelium discoide	7.83e-09
29	47	5.4	339	7	C23862 Dictyostelium discoide	7.83e-09
30	47	5.4	396	14	AA980543 ua42e08.r1 Soares mous	7.83e-09
31	47	5.4	405	18	AI180802 ub76e02.r1 Soares mous	7.83e-09
32	47	5.4	476	12	AA747432 nx88f03.s1 NCI CGAP GC	7.83e-09
33	47	5.4	580	14	C91340 Dictyostelium discoide	7.83e-09
34	47	5.4	614	14	C91555 Dictyostelium discoide	7.83e-09
35	46	5.3	183	14	C93036 Dictyostelium discoide	7.83e-09
36	46	5.3	200	15	C93770 Dictyostelium discoide	7.83e-09
37	46	5.3	224	11	C84161 Dictyostelium discoide	7.83e-09
38	46	5.3	247	12	AA754458 97SN1784 Rice Immature	2.72e-08
39	46	5.3	260	15	AI049844 an21e05.x1 Gessler Wil	2.72e-08
40	46	5.3	341	12	AA748451 ca56e01.s1 NCI CGAP GC	2.72e-08
41	46	5.3	378	18	AI146542 qb45a12.x1 NCI CGAP Br	2.72e-08
42	46	5.3	437	15	C94484 Dictyostelium discoide	2.72e-08
43	46	5.3	450	14	C89917 Dictyostelium discoide	2.72e-08
44	46	5.3	459	24	AA266894 mz98h12.r1 Soares mous	2.72e-08
45	46	5.3	621	26	B21581 T25117-Sp6 TAMU Arabid	2.72e-08

ALIGNMENTS

RESULT	1	AA311108	496 bp	mRNA	EST	19-APR-1997
LOCUS		EST181872	Jurkat T-cells	V Homo sapiens	CDNA 5' end, mRNA sequence.	
DEFINITION		AA311108				
ACCESSION		AA311108				
NID		G1963435				
KEYWORDS		EST.				
SOURCE		human.				
ORGANISM		Homo sapiens				
REFERENCE						
AUTHORS						

1 (bases 1 to 496)
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S., Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palancas,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dinkar,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Raymond,L.L., Kunsch,C., Hungjun,J., Li,H., Weisner,P.S., Olsen,H., Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.N. and Venter,J.C.
Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
JOURNAL - Nature 377 (6547 Suppl), 3-174 (1995)
REMARKS - 96026380
COMMENT - Other ESTs: THC124424
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018599056

Fax: 3018599423

Email: arkerlavetigr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (<http://www.tigr.org/tdb/hgl/hgl.html>)

Seq primer: M13 Reverse

Location/Qualifiers

FEATURES

source

1. .496

/organism="Homo sapiens"

/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:

XhoI"

/db_xref="ATCC (inhost):157086"

/db_xref="taxon:9606"

/clone_lib="Jurkat T-cells v"

/cell_type="T-lymphocyte"

<1. .496

98 a 143 c 151 g 103 t 1 others

Query Match 53.5%; Score 463; DB 24; Length 496;

Best Local Similarity 99.2%; Pred. No. 2.06e-287;

Matches 482; Conservative 0; Mismatches 1; Indels 3; Gaps 3;

Db 5 TTTGTGGAGGAGCAGAGATACCCAGCTGGACATCTTCTGCTGATGAGCCCCAGG 64

Qy 1 tttgtggaggagcagagatgccagctggacatcttctctgtatgagcccccagg 60

Db 65 CTGGAGGTGCTCTGCTCATCTCTTCCAGGGTCTCTCGCTGGGAGGTCTCATCA 124

Qy 61 ctggagggtgccctgtctcatctcttcccagggtctctgcctggcagggtcatcata 120

Db 125 GTACGGGAGCTGCTCTGCAAGCCGAGCAATTTACTGTGAGCTGAGGGACCAAGCT 184

Qy 121 gtacggggagctggtcttgcaggagccagcagcattttactgtgagcctggggaccaggct 180

Db 185 GCCATGCTCTGTGACACTAGGGCTCTTTCGACAGACACTCTGGCCTGTCTCTCC 244

Qy 181 gccatgctctgtgacactagggcctcttcgcagacagaaactctggcctggatctcc 240

Db 245 CGCTGGGGCAGAGAAACTGATCTAGCCCTCTCTCTCTTTACCCCGCAGAGATCTTT 304

Qy 241 cgctggggcagaagaactgactcagccctctctcttttaccgccagagattcttt 300

Db 305 GAGGTGCTGCTCTGTTCAGAGGGAGGCTGAAGCTGCGCTCAATGGCAGGGGCTG 364

Qy 301 gaggtgctgctctgttccaggaggagggctggaagctggcctcaatggcagggtcg 360

Db 365 GGGGACACAGCATGAACCCAGAGGCCCTGGAGCAGCTCGGGAGCTCCGGATCAGTGA 424

Qy 361 ggggacacagcatgaaccagcagccctggagcagctcgggagctccggatcagtgga 420

Db 425 AGTGTCCAGCTTACTGTCTCAGCTCTCTGA-GGATGGTTCCAGGGAATACCCGNCAGAAA 483

Qy 421 agtgtccagctactgtgtccactctgagaggatggttccagg-aaataccg-cagaaa 478

Db 484 ACAAGA 489

Qy 479 acaaga 484

RESULT 2 H30148 411 bp mRNA EST 16-AUG-1995

LOCUS Y059h05.r1 Homo sapiens cDNA clone 182265 5'.

ACCESSION H30148

NID 9901058

KEYWORDS EST.

SOURCE human clone-182265 library-Soares breast 3DNBst vector-pt73D

(Pharmacia) with a modified polylinker host-DH10B (ampicillin

resistant) primer-M13RPI Rsite1-Not I Rsite2-Eco RI Adult human.

1st strand cDNA was primed with a Not I - oligo(dT) primer [5'

TGTACCAATCTGAAGTGGGAGCGCGCCCTTTTTTTTTTTTTTTT 3'],

double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 20. Library constructed by Bento Soares and M.Fatima Bonaldo.

ORGANISM

Homo sapiens

Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE

AUTHORS

Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Fan,F., Trevisan,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.

The WashU-Merck EST Project

Unpublished (1995)

COMMENT

Contact: Wilson RK

WashU-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 850L, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

High quality sequence stops: 264

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES

source

Location/Qualifiers

1. .411

/organism="Homo sapiens"

/clone="182265"

BASE COUNT 90 a 113 c 121 g 79 t 8 others

ORIGIN

Query Match 38.8%; Score 336; DB 16; Length 411;

Best Local Similarity 95.7%; Pred. No. 2.64e-199;

Matches 360; Conservative 0; Mismatches 13; Indels 3; Gaps 3;

Db 1 CAGGGTCTCTNACAGGGCAGGTCATCATAGTAGGGGGANTGGTCTTNCAGAGCCGAG 60

Qy 91 cagggtctctcgctggcaggtcatcatagtagcgggactggtcttgcaggagcgaag 150

Db 61 CATTTACTGTGAGCTGAGGACGACGCTNCCATGCTCTGTGACACTCAGGCCCTCC 120

Qy 151 catttactgtgagctgaggaccaggtgcccactgctctgtgacactcagggcctcc 210

Db 121 TTGCGACAGAACTCTGGCTGGATCTCCGCTGGGGGCGAGAAACTGATCTCAGCC 180

Qy 211 ttgcagacagaaactctggctgattccctccgtggggcagagaaactgattcagcc 270

Db 181 CCCTTCTCTTTTACCCNCAGAGATCTTTGAGGTGCTGCTCTCTGTTCCAGGAGGG 240

Qy 271 cccttctcttttaccgccagagattttgaggtgctgctctgttccagggaggg 330

Db 241 CTGAAGCTGCGCTCAATGGGCGAGGGCTGGGGCCGACCAAGCAATGAACAGAGCCC 300

Qy 331 ctgaagctgctgctcaatggcaggggctgggggcccacca-gca-tgaaccagcagggccc 388

Db 301 TGGAGCAGCTGCGGNGACTCCGGATCAGTGGAAAGTATCCAGCTTACTGTCTCCACTCT 360

Qy 389 tggagcagctgcgaggagctcccgatcagtggaagtgtccagctactgtgtccactcct 448

Db 361 NA-GGATGGTTCAGG 375

Qy 449 gaaggtggttccagg 464

RESULT 3 N27468 422 bp mRNA EST 16-AUG-1995

LOCUS Y059h05.s1 Homo sapiens cDNA clone 182265 3'.

DEFINITION

Db 359 TCCTGCAGCAGTCCTCCACGGGCCCTGCTGGTGTTCANGCTGGTGCGGCCCCCAGNCCTTCCCAT 418
||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Cp 404 tccgcagctgctccagg-ccgtcgtgttcattgctgttg-ccccacgccctggccat 347

Db 419 T 419
I
Cp 346 t 346

RESULT 4
LOCUS RA349100 227 bp mRNA EST 21-APR-1997
DEFINITION EST55685 Infant adrenal gland II Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION AA349100
NID G2001358
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondria eukaryotes; Metazoa; Chordata;
Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea;
Homo.

REFERENCE 1 (bases 1 to 227)
ADAMS,M.D., KERLAVAGE,A.R., FLEISCHMANN,R.D., FULDNER,R.A.,
BULT,C.J., LEE,N.H., KIRKNESS,E.F., WEINSTOCK,K.G., GOCCAYNE,J.D.,
WHITE,O., SUTTON,G., BLAKE,J.A., BRANDON,R.C., MAN-WAI-C.,
CLAYTON,R.A., CLINE,T.R., COTTON,M.D., EARLE-EUCHES,J., FINE,L.D.,
FITZGERALD,L.M., FITZHUGH,W.M., FRITCHMAN,J.L., GEORGHAGEN,N.S.,
GLOCKE,A., GHELM,C.L., HANNA,M.C., HEDBLUM,E., HINKLE,P.S., JR.,
KEILLY,J.M., KELLEY,J.C., LIU,L.-I., MARMAROS,S.M., MERRICK,J.M.,
MORENO-PALANQUES,R.F., McDONALD,L.A., NGUYEN,D.T., PELLIGRINO,S.M.,
PHILLIPS,C.A., RYDER,S.E., SCOTT,J.L., SAUDEK,D.M., SHIRLEY,R.,
SMALL,K.V., SPRIGGS,T.A., UTTERBACK,T.R., WEIDMAN,J.F., LI,Y.,
BEDNARIK,D.P., CAO,L., CEPEDA,M.A., COLEMAN,T.A., COLLINS,E.J.,
DIMKE,D., FENG,D.-F., FERIE,A., FISCHER,C., HASTINGS,G.A.,
HE,W.W., HU,J.S., GREENE,J.M., GRUBER,J., HUDSON,P., KIM,A.K.,
KOZAK,D.L., KUNSCHE,C., HUNGJUN,J., LI,H., MEISSNER,P.S., OLSEN,H.,
RAYMOND,I., WEI,Y.F., WING,J., XU,C., YU,G.L., RUBEN,S.M.,
DILLON,P.J., FANNON,M.R., ROSEN,C.A., HASSELTINE,W.A., FIELDS,C.,
FRASER,C.M. and VENTER,J.C.

TITLE Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence

JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)

MEDLINE 96026280

COMMENT Other_ESRs: THCl24424
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlavet@tigr.org
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/cdb/hgi/hgi.html>)
Seq primer: M13 Reverse.
Location/Qualifiers
1..227
/organism="Homo sapiens"
/note="Organ: adrenal gland; Vector: pbluescript SK-";
Site_1: EcoRI; Site_2: XhoI"
/db_xref="ATCC (inhost):150218"
/db_xref="taxon:9606"
/clone_lib="Infant adrenal gland II"
/sex="male"
/dev_stage="infant, 16 wks"
<1..>227

BASE COUNT 42 a 72 c 60 g 51 t 2 others

ORIGIN

Query Match 25.5%; Score 221; DB 25; Length 227;
Best Local Similarity 98.2%; Fred. No. 1.44e-120;
Matches 223; Conservative 0; Mismatches 4; Indels 0 Gaps 0;

```

Db 1 AGAGTACCCAGCTGGACATCCTTTCTGCTGATGAGCCCCAGGCTGGAGTGCCTGCTC 60
  |||
QY 18 agagtagccagctggacatccttctctgctgatgagccccagctggaggtgccccgctc 77
  |||
Db 61 ACATGCTCTCCCGAGGCTCTCCGCTGGGCGAGTCAATAGTACGGGAGCTGCTCTT 120
  |||
QY 78 acatgctctcccgaggtctctgctgctggcgaggtcctcatcagtcagggagctgtt 137
  |||
Db 121 GCAGAGCCGAGCAATTTACTGTAAAGCTTGAGGACACAGGCTNCCCATGCTCTTAC 180
  |||
QY 138 gcaagagccgaagcatcttactgtgagctgagggagccaggtgcccctgctctgtac 197
  |||
Db 181 ACTCAGGGCTCTCTCCGAGACAGAACATGCGCTGGATCTCCCGCT 227
  |||
QY 198 actcagggctctctccgagacagaaactctggcctggatctccccgt 244
  |||

RESULT 5 AA020729 660 bp mRNA EST 30-JAN-1997
LOCUS ze63all.r1 Soares retina N2b4HR Homo sapiens cDNA clone 363644 5'
DEFINITION Similar to contains LI.tl LI repetitive element ;, mRNA sequence.
ACCESSION AA020729
NID 91484512
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 660)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le.M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 2180 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 444.
location/Qualifiers
1. .660
/organism="Homo sapiens"
/note="Organ: eye; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). The retinas were obtained from a 55 year old
Caucasian and total cellular poly(A)+ RNA was extracted 6
hrs after their removal. The retina RNA was kindly
provided by Roderick R. McInnes M.D. Ph.D. from the
University of Toronto. Library constructed by Bento
Soares and M.Fatima Bonaldo."
/db_xref="taxon:9606"
/clone_lib="Soares retina N2b4HR"
/sex="male"
/tissue_type="retina"
/dev_stage="55 year old"
/lab_host="DH10B (ampicillin resistant)"
<1. .>660
198 a 147 c 168 g 142 t 5 others

FEATURES
source
/organism="Homo sapiens"
/note="Organ: eye; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). The retinas were obtained from a 55 year old
Caucasian and total cellular poly(A)+ RNA was extracted 6
hrs after their removal. The retina RNA was kindly
provided by Roderick R. McInnes M.D. Ph.D. from the
University of Toronto. Library constructed by Bento
Soares and M.Fatima Bonaldo."
/db_xref="taxon:9606"
/clone_lib="Soares retina N2b4HR"
/sex="male"
/tissue_type="retina"
/dev_stage="55 year old"
/lab_host="DH10B (ampicillin resistant)"
<1. .>660
198 a 147 c 168 g 142 t 5 others

BASE COUNT
mRNA

```

```

ORIGIN
Query Match 13.9%; Score 120; DB 22; Length 660;
Best local similarity 96.2%; Pred. No. 2.17e-53;
Matches 150; Conservative 0; Mismatches 1; Indels 5; Gaps 5;

Db 1 CTGAAGCTGGCGCTCAATGGCAGGGCTGGGGGC-ACCAGCATGACACGAGGCCCTG 59
  |||
QY 331 ctgaagctggcgctcaatggcgagggctggggccaccagcatgaaccagcggccctg 390
  |||
Db 60 GAGCAGCTCGGNGA-CTCCGGATCAGTGAAGTGTCCAGCTCTACTGTGTCTCTCTGA 118
  |||
QY 391 gagcagctcgggagctccggatcagtggaagtgctcagctctactgtgtccactctga 450
  |||
Db 119 -GGATGTTCCAGGGAATACCGCCAGAAACAAAGA 153
  |||
QY 451 aggatggttccagg-aaataccgc-agaaaacaaga 484
  |||

RESULT 6 AA01029 434 bp mRNA EST 29-NOV-1996
LOCUS ze46b12.r1 Soares retina N2b4HR Homo sapiens cDNA clone 362015 5'
DEFINITION similar to contains Alu repetitive element; contains LI.tl LI
repetitive element ;, mRNA sequence.
ACCESSION AA01029
NID 91437096
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 434)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le.M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
Wilson,R.
WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 2264 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 345.
location/Qualifiers
1. .434
/organism="Homo sapiens"
/note="Organ: eye; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). The retinas were obtained from a 55 year old
Caucasian and total cellular poly(A)+ RNA was extracted 6
hrs after their removal. The retina RNA was kindly
provided by Roderick R. McInnes M.D. Ph.D. from the
University of Toronto. Library constructed by Bento
Soares and M.Fatima Bonaldo."
/db_xref="taxon:9606"
/clone_lib="Soares retina N2b4HR"
/sex="male"
/tissue_type="retina"

```

```

/dev_stage="55 year old"
/lab_host="DH10B (ampicillin resistant)"
<1. ->434
BASE COUNT      128 a  104 c  104 g  94 t      4 others
ORIGIN

Query Match      9.2%; Score 80; DB 22; Length 434;
Best Local Similarity 94.5%; Pred. No. 4.14e-28;
Matches 104; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

Db 1 ACCAGCAGCCCTGGAGAGCTGGGNGA-CYCCGGATCAGTGAAGTNPCCAGCTCTACT 59
|||||
Qy 377 accagcagccctggagcagctggcgagctccgagctcagtgatggaagtgccagctctact 436
|||||

Db 60 GTGTCCTCTCTGA-GGATGCTTCAGGGAATACCGCCAGCAAAACAAGA 108
|||||
Qy 437 gtgtccactcctgaaggtggttcagg-aataccgc-agaaaacaaga 484
|||||

RESULT 7
LOCUS      AA754459      252 bp      mRNA      EST      20-JAN-1998
DEFINITION 97SN1787 Rice Immature Seed Lambda ZAPII CDNA Library Oryza sativa
ACCESSION  AA754459
NID         92801165
KEYWORDS   EST.
SOURCE     rice.
ORGANISM   Oryza sativa
            Eukaryotae; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; seed plants; Magnoliophyta; Liliopsida;
            Poales; Poaceae; Oryza.
REFERENCE  1 (bases 1 to 252)
AUTHORS   Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,
            Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
            Lee,M.C. and Eun,M.Y.
TITLE     Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
JOURNAL   Unpublished (1998)
COMMENT   Contact: Eun M.Y.
            Department of Cytogenetics
            National Inst. of Agri. Sci. and Tech, RDA
            Suwon, Kyunggido, Korea
            Tel: 82 331 290 0301
            Fax: 82 331 290 0307
            Email: myeun@sun20.asti.re.kr
            Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
            University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
            Seq primer: M13 Reverse Primer.
            Location/Qualifiers
                1. .252
                /organism="Oryza sativa"
                /cultivar="Milyang23"
                /note="Vector: pBluescript SK(+); Site.1: EcoRI; Site.2:
                XhoI; Directional CDNA library inserted into lambda ZAPII
                vector at 5' end with EcoRI and 3' end with Xho I site."
                /db_xref="taxon:4530"
                /clone="97SN1787"
                /clone_lib="Rice Immature Seed Lambda ZAPII CDNA Library"
                /tissue_type="Immature Seed"
                /dev_stage="5 days after pollination"
                /lab_host="E. coli SOLR"

BASE COUNT      5 a  21 c  12 g  35 t  179 others
ORIGIN

Query Match      7.1%; Score 61; DB 12; Length 252;
Best Local Similarity 14.3%; Pred. No. 9.89e-17;
Matches 32; Conservative 110; Mismatches 79; Indels 3; Gaps 3;

Db 13 CCCAAWTTTSYBCHGNBVCVASHGNYMSVHN-CTBGTCTCCKCNVKNVSTWGTNVNB 71
|||||
Qy 534 cctgaacaccagcacatcaggcctgttcacctctgggttcacagactgagctacag 593
|||||
Db 72 NVSGDWHYWBYSNTKVDVGNETRC-SRWBVTMAHYDHTNCBYYNNNDYHWHBMBZB 130
|||||

```

```

Qy 594 ggccttg-ggcctgagggaagcacaagagtgcaaggttctctgaacttcgaccttc 652
|||||
Db 131 BFGCTCTTACWCBHNTKCTAGWHTSTWYDKSNTTWGTSYDKSHGWCBBVYK 190
|||||
Qy 653 ctccaccaggagcctggatgctccatctgcttcagggcctgactgactcacag 712
|||||
Db 191 HNKVSTTRATRSYTCVRKYCVMMTKVKVKKVYVBBGCHBTD 234
|||||
Qy 713 aggcgaagtgttagactaacaagatactccaaaatacaatgg 756
|||||

RESULT 8
LOCUS      AA754459      252 bp      mRNA      EST      20-JAN-1998
DEFINITION 97SN1787 Rice Immature Seed Lambda ZAPII CDNA Library Oryza sativa
ACCESSION  AA754459
NID         92801165
KEYWORDS   EST.
SOURCE     rice.
ORGANISM   Oryza sativa
            Eukaryotae; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; seed plants; Magnoliophyta; Liliopsida;
            Poales; Poaceae; Oryza.
REFERENCE  1 (bases 1 to 252)
AUTHORS   Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,
            Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
            Lee,M.C. and Eun,M.Y.
TITLE     Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
JOURNAL   Unpublished (1998)
COMMENT   Contact: Eun M.Y.
            Department of Cytogenetics
            National Inst. of Agri. Sci. and Tech, RDA
            Suwon, Kyunggido, Korea
            Tel: 82 331 290 0301
            Fax: 82 331 290 0307
            Email: myeun@sun20.asti.re.kr
            Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
            University, Yongin, Korea. 449-728 bhnahm@bioserver.myongji.ac.kr
            Seq primer: M13 Reverse Primer.
            Location/Qualifiers
                1. .252
                /organism="Oryza sativa"
                /cultivar="Milyang23"
                /note="Vector: pBluescript SK(+); Site.1: EcoRI; Site.2:
                XhoI; Directional CDNA library inserted into lambda ZAPII
                vector at 5' end with EcoRI and 3' end with Xho I site."
                /db_xref="taxon:4530"
                /clone="97SN1787"
                /clone_lib="Rice Immature Seed Lambda ZAPII CDNA Library"
                /tissue_type="Immature Seed"
                /dev_stage="5 days after pollination"
                /lab_host="E. coli SOLR"

BASE COUNT      5 a  21 c  12 g  35 t  179 others
ORIGIN

Query Match      6.4%; Score 55; DB 12; Length 252;
Best Local Similarity 10.0%; Pred. No. 2.79e-13;
Matches 21; Conservative 109; Mismatches 78; Indels 2; Gaps 2;

Db 41 MSVHNTCBRTGTHCDCKCNVKNVSTWGTNVBNVSGDWHYBNTKVDVGNHTRCSEWB 100
|||||
Qy 771 accacattcttaagcattgctatttggagctattctttagctctacacacttgcct-t 713
|||||
Db 101 VTRMAHYDHTNCBYYNNNDYHWHBMBYBFTGCMCTCTACWCBHNTKCTASGNWSTNY 160
|||||
Qy 712 cctgtagtcagtcagccctgaaggcagatggagccatataccacagctcctggtgag 653
|||||
Db 161 DVKSSTNWGV-TBSYDKSMHGWCSBBYHKTSTTRATRSYTCVRKYCVMMTKVKV 219
|||||
Qy 652 gaaggtgcagagtcaggaacactttgctctcttgcctctccctcagcccaagctcc 593
|||||

```


[illegible]

```

Matches      54;  Conservative      0;  Mismatches      6;  Indels      0;  Gaps      0
Db 159 AAAAAAAAAACAATTTATTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 218
    II III III III III III III III III III III III III III
Qy 806 aataataaaggtattattataaaaaaataaaaaaataaaaaaataaaaaa 865

Search completed: Tue Dec 15 18:57:34 1998
Job time : 1612 secs.

```

(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_unn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 15 18:00:04 1998; MasPar time 1259.73 Seconds
Tabular output not generated. 1513.252 Million cell updates/sec

Title: >US-09-109-864-1
Description: (1-865) from US09109864.seq
Perfect Score: 865
N.A. Sequence: 1 tttgtgaggcgcagcagaga.....aaaaaaaaaaaaaaaaaaaaa 865
Comp: aacacctccgtgctctct.....tttttttttttttttttt

Scoring table: TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 567134 seqs, 1101898692 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: emb155
1:em_ba 2:em_fun 3:em_htg 4:em_hum1 5:em_hum2 6:em_in
7:em_om 8:em_or 9:em_ov 10:em_pat 11:em_ph 12:em_pl
13:em_ro 14:em_vl

Database: genbank107
15:gb_bal 16:gb_ba2 17:gb_htg 18:gb_in 19:gb_om 20:gb_ov
21:gb_pat 22:gb_ph 23:gb_pl1 24:gb_pl2 25:gb_pr1
26:gb_pr2 27:gb_pr3 28:gb_ro 29:gb_st 30:gb_sts 31:gb_sy
32:gb_un 33:gb_vl

Statistics: Mean 11.899; Variance 15.521; scale 0.767

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
c 1	477	55.1	35414	26	Human Chromosome 11 Co	2.20e-137
2	54	6.2	2982	23	Dictyostellu discoideu	3.81e-04
3	48	5.5	1927	25	H.sapiens mRNA for 50	1.50e-02
4	48	5.5	2787	18	P.falciparum gene for	1.50e-02
5	48	5.5	24649	17	*** SEQUENCING IN PROG	1.50e-02
6	47	5.4	144	21	Sequence 26 from patent	2.74e-02
7	47	5.4	766	21	Sequence 1 from patent	2.74e-02
8	47	5.4	2606	23	Dictyostellum discoide	2.74e-02
9	47	5.4	10772	18	Drosophila melanogaste	2.74e-02
10	46	5.3	460	18	Lucilia cuprina cuticl	4.97e-02
11	46	5.3	1744	23	A.thaliana mRNA for th	4.97e-02
12	46	5.3	1999	28	Mus musculus putative	4.97e-02
13	46	5.3	3795	19	Bos taurus Fanconi ane	4.97e-02

c 14	5.3	10772	18	AF012089	Drosophila melanogaste	4.97e-02
15	5.2	472	33	BCU45456	Hog cholera virus stra	9.01e-02
c 16	5.2	542	18	PFNRSUB	P.falciparum non-repet	9.01e-02
17	5.2	553	18	BAY12278	H.armigera mRNA for pu	9.01e-02
18	5.2	1438	18	DMU86867	Drosophila melanogaste	9.01e-02
19	5.2	1606	18	DDU41222	Dictyostellum discoide	9.01e-02
20	5.2	1690	28	RNU76551	Rattus norvegicus muc1	9.01e-02
21	5.2	2253	18	AF032396	Hyalophora cecropia ar	9.01e-02
22	5.2	2794	18	DDU48271	Dictyostellum discoide	9.01e-02
23	5.2	3437	21	AS2184	Sequence 17 from Paten	9.01e-02
24	5.2	3437	23	SOPULSPO	S.oleracea L. mRNA for	9.01e-02
25	5.2	4826	26	HSRANBP5	H.sapiens mRNA for Ran	9.01e-02
c 26	5.2	10003	18	PFSC04009	Plasmodium falciparum	9.01e-02
c 27	5.2	13445	18	PFSC04059	Plasmodium falciparum	9.01e-02
c 28	5.2	76848	17	PFMALIP2	Plasmodium falciparum	9.01e-02
c 29	5.2	129169	17	AC005506	*** SEQUENCING IN PROG	9.01e-02
c 30	5.1	829	26	BSAJ3352	Homo sapiens mRNA for	1.63e-01
31	5.1	1086	18	HAY12282	B.armigera mRNA for pu	1.63e-01
c 32	5.1	2483	23	DD113A	Dictyostellum discoide	1.63e-01
c 33	5.1	3363	18	CEMCEIA	C.elegans mRNA for myo	1.63e-01
c 34	5.1	7218	21	I66494	Sequence 14 from Paten	1.63e-01
35	5.1	8574	18	AF057019	Dictyostellum discoide	1.63e-01
c 36	5.1	151018	17	PFMALIP6	Plasmodium falciparum	1.63e-01
c 37	5.0	340	21	I22446	Sequence 27 from Paten	2.92e-01
c 38	5.0	350	21	I22433	Sequence 14 from Paten	2.92e-01
39	5.0	1013	28	RNU75932	Rattus norvegicus CAMP	2.92e-01
40	5.0	1136	21	I08009	Sequence 3 from Patent	2.92e-01
41	5.0	1558	21	P03671	DNA sequence coding fo	2.92e-01
42	5.0	1558	21	P66401	Sequence 7 from patent	2.92e-01
43	5.0	1718	26	HSAJ3198	Homo sapiens mRNA for	2.92e-01
44	5.0	2874	18	S52651	RNA polymerase II lary	2.92e-01
45	5.0	7218	21	I66494	Sequence 14 from Paten	2.92e-01

ALIGNMENTS

RESULT 1	U73641	35414 bp	DNA	PRI	25-JUL-1997
LOCUS	Human Chromosome 11	Cosmid cSRL187d6	complete sequence.		
DEFINITION	U73641				
ACCESSION	U73641				
NID	G2281065				
KEYWORDS	HTG.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 35414) Evans, G.A., Bradbury, P., Brignac, S., Bumeister, R., Burbee, D., Davie, J., Davies, C.J., Davis, C., English, C., Fondon, T., Franklin, T.L., Garner, H.R., Gordon, M., Gotway, G., Grant, O., Hahner, L., Harris, J., Hinson, S., Megarity, C., Narayanaswamy, U., Newton, J., O'Brien, K., Oliver, P., Patel, P., Probst, S., Rayner, S., Schagenham, J., Schilling, P., Schultz, R., Syed, M., Valenzuela, D., Ward, T. and Wilson, R.				
TITLE	Template				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 35414) Evans, G.A., Bradbury, P., Brignac, S., Bumeister, R., Burbee, D., Davie, J., Davies, C.J., Davis, C., English, C., Fondon, T., Franklin, T.L., Garner, H.R., Gordon, M., Gotway, G., Grant, O., Hahner, L., Harris, J., Hinson, S., Megarity, C., Narayanaswamy, U., Newton, J., O'Brien, K., Oliver, P., Patel, P., Probst, S., Rayner, S., Schagenham, J., Schilling, P., Schultz, R., Syed, M., Valenzuela, D., Ward, T. and Wilson, R.				
AUTHORS	Submitted (25-JUL-1997) Genome Science and Technology Center, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA On Jul 25, 1997 this sequence version replaced gi:1737196. Location/Qualifiers 1. .35414 /organism="Homo sapiens"				
TITLE	Direct Submission				
JOURNAL					
COMMENT					
FEATURES					
source					


```
/db_xref="taxon:44689"
/tissue_type="germinating spore"
984..1253
/gene="CelB"
/number=1
984..2881
/gene="CelB"
join(1196..1253,1341..2881)
/gene="CelB"
/note="spore germination-specific protein; formerly called
270-11"
/codon_start=1
/product="cellulose binding protein"
/db_xref="PID:g167881"
/translation="MKNIYSLFLFALLSATFANNAFIVHNSDSISKLLGQIGDTI
SFYSDSHDVKSDSDSVSVFSGSLNPGIFKVLTKEGNIEFTSSYDEGLSATI
VSSGGQIPITSTSTTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST
SSGTGSGTSAVTTSSSSNDSDGSGVGTSTTSPAITSSGSIIDPTSPPTDSSS
NSGGYSSSSSIENGVECLLITQDAFDSMTYDNLITVYOVNLTGSLVESVILTP
NDNSLIHTWELVDGSLILPYRKAGPINPEETIIFGYISRMSTDTVPALSPICSD
SSSPPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
SEIETPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
INGEDVBIQVEATIVNQGSTPISSFNFYSDAEQIWSVEKGTNTYKLPFSFSTIPVG
GSTFGYIVKSAELSDLEGVQYC"
1254..1340
/gene="CelB"
1341..2881
/gene="CelB"
/number=2
BASE COUNT 1168 a 417 c 306 g 1091 t
ORIGIN

Query Match 6.2%; Score 54; DB 23; Length 2982;
Best Local Similarity 79.3%; Pred. No. 3.81e-04;
Matches 73; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Db 349 TTTTATTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAA 408
QY 774 tttattctttattttatttttttttttttttttttttttttttttttttttt 833
Db 409 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 440
QY 834 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 865

RESULT 3 HSEPM650 1927 bp RNA PRI 30-JUN-1993
LOCUS H.sapiens mRNA for 50 kDa erythrocyte plasma membrane glycoprotein.
DEFINITION X64594 S46252
ACCESSION g31194
NID glycoprotein; plasma glycoprotein.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM
REFERENCE Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
AUTHORS Vertebrata; Euthera; Primates; Catarrhini; Homidae; Homo.
TITLE 1 (bases 1 to 1927)
JOURNAL Direct Submission
AUTHORS Ridgwell, K.
SUBMITTED (23-MAR-1992) K. Ridgwell, University of Bristol, Dept of
JOURNAL Biochemistry, School of Medical Sciences, University Walk, Bristol
REFERENCE BS 17D, UK
AUTHORS 2 (bases 1 to 1927)
Ridgwell, K., Spurr, N.K., Laguda, B., MacGeoch, C., Avent, N.D. and
Tanner, M.J.
TITLE Isolation of cDNA clones for a 50 kDa glycoprotein of the human
erythrocyte membrane associated with Rh (rhesus) blood-group
antigen expression
JOURNAL Biochem. J. 287 (Pt 1), 223-228 (1992)
MEDLINE 93038558
FEATURES Location/Qualifiers
source 1..1927
/organism="Homo sapiens"
/db_xref="taxon:9606"
```

```
/dev_stage="foetal"
/tissue_type="bone marrow, liver"
/chromosome="6"
/map="p21-qter"
28..1257
CDS
/codon_start=1
/product="50 kDa erythrocyte plasma membrane glycoprotein"
/db_xref="PID:g31195"
/db_xref="SWISS-PROT:Q02094"
/translation="NRFTPLMAIVLEIAMIVLFGVLEIYETDQVLEQINIKPTDM
GIEFFLYPLFDVHMVIFWFGFGLATFLKKGYSVGINLLVAALGLONGTIVQGLQ
SOGKFNIGIKNMADPSAATVLSFGAVLGKTSPTOMLINTILEIYFAHNEVLVS
EIFKASDIGASNTIHAFGAYFGLAVAGILYRSLKRGHENSEATYSDLFAMIGILFL
MWFPSFSAIAEPGDKOCRAIVDTYIFSLAACVLTAFAFSSLSVEHGRKLNMYHQNAT
LAGGVAVGTCDMAIHPFGSMITIGSIAGHSVLYGKFLPTFTTLRLIHDTGCVNHL
GLPGVYVGLAGIVAVMAGASNTSMAGAAALGSSIGTAVVGGLMTGLILKLPWGQPS
DONCVDSSVYMKVPTKTR"
136..138
/misc_feature
/standard_name="glycosylation site"
307..309
/misc_feature
/standard_name="glycosylation site"
390..392
/misc_feature
/standard_name="glycosylation site"
751
/misc_feature
/note="ASN/ASP polymorphism"
BASE COUNT 563 a 392 c 411 g 561 t
ORIGIN

Query Match 5.5%; Score 48; DB 25; Length 1927;
Best Local Similarity 87.5%; Pred. No. 1.50e-02;
Matches 56; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 1864 GTCCATAATTAATATTGTATATAAATAAAAAAAAAAAAAAAAAAAAAA 1923
QY 802 gtcaataataataaagttatttttttttttttttttttttttttttttt 861
Db 1924 AAAA 1927
QY 862 aaaa 865

RESULT 4
LOCUS PFSTARP 2787 bp DNA INV 11-MAY-1995
DEFINITION P.falci-parum gene for STARP antigen.
ACCESSION Z26314
NID G499324
KEYWORDS STARP antigen.
SOURCE malaria parasite.
ORGANISM Plasmodium falciparum
REFERENCE Eukaryotes; mitochondrial eukaryotes; Alveolata; Apicomplexa;
AUTHORS Haemosporida; Plasmodium.
TITLE 1 (bases 1 to 2787)
JOURNAL Fidock, D.A., Bottius, E., Brahimi, K., Moelans, I., I.M.D., Alkawa, M.,
MEDLINE Konings, R.N.H., Certa, U., Olafsson, P., Kaidoh, T., Asavanich, A.,
AUTHORS Guerin-Marchand, C. and Drullhe, P.
TITLE Cloning and characterization of a novel Plasmodium falciparum
JOURNAL sporozoite surface antigen, STARP
REFERENCE Mcl. Biochem. Parasitol. 64 (2), 219-232 (1994)
MEDLINE 95021499
AUTHORS 2 (bases 1 to 2787)
Fidock, D.A.
TITLE Direct Submission
JOURNAL Submitted (16-SEP-1993) David A. Fidock, Laboratory of bio-medical
parassitology, Pasteur, Institute, 25, rue du Dr. Roux, Paris,
PARIS, 75724 CEDEX 15, France
FEATURES Location/Qualifiers
source 1..2787
/organism="Plasmodium falciparum"
/strain="T9/96"
/db_xref="taxon:5833"
/clone="DG21"
/dev_stage="Sporozoite, liver stage, erythrocytic ring
stage"
```

```

/cell_line="T9/96 blood stage culture"
/clone_lib="Lambda gtl1"
/chromosome="8"
1..806
/number=1
join(1..806,982..2787)
/product="STARP antigen"
join(735..806,982..2724)
/codon_start=1
/product="STARP antigen"
/db_xref="PID:9499325"
/translation="MIHIFPKTAFTLSLITWTLTLLSNKLNKCNFYNNNNLSTYVIKH
NRFLSEYQSLGGGYSAAKLKLVSKSGTNTVNTYNSNTNTNNIPSSSTYVNTNR
LAANNSTTSIKVDNNKNIKILGNNSTTNTSTNTSNTATKVTENVITNQILITG
NNNTNTSNDNSTNTLNDNTSTTKKLNDNTNTQNLTTSTNTTSTVSTNNNT
KPDNNNTDIKSTNTGTETDKNTDKIDKATDNNNTITQNLTTSTNTTSTVSTNNKTN
VISTDNTSTNTISTDNDNADTILDNDNTDILTDNNNTDILSTDNDNADKATDNN
NNTWKATDNNNTKLIISPDNNNTKTSIDNNNTNNKATDNNNTKTSIDNNNTKLIIS
TDNNNTKTSIDNNNTKTSIDNNNTNTISTDNNNTNTQYFANNYNETTSDDELNK
DSCDYSEKENIKSINAYLDKLDLETVRKHSDISTCIERKNRPNQITHLNNLNKM
YNIKFIIVIIYIAFNWSEVIYKVGKILAFALYMLIN"
807..981
/number=1
982..2787
/number=2
BASE COUNT 1411 a 306 c 210 g 860 t
ORIGIN
    Query Match      5.5%; Score 48; DB 18; Length 2787;
    Best Local Similarity 77.3%; Pred. No. 1.50e-02;
    Matches 68; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Db  444 TTCATTATTCCTTTATTTTAAACCAATAATAATATATATATAGATTAAAAAAA 503
Qy  778 tctttattattattatttgggtcaataataataaagggtattttattcaaaaaa 837
      ||| ||||| || | ||| ||||| ||||| ||||| ||| ||||| |||||
Db  504 AAAAAAATAAAAAAATAAAAAA 531
Qy  838 aaaaaaataaaaaaataaaaaa 865
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
LOCUS AC005308 244649 bp DNA HTG 20-AUG-1998
DEFINITION *** SEQUENCING IN PROGRESS *** Plasmodium falciparum 3D7 chromosome
12 PFVAC492 genomic sequence; HTGS phase 1, 36 unordered pieces.
ACCESSION AC005308
NID 93445189
KEYWORDS HTG; HTGS_PHASE1.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 244649)
AUTHORS Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
TITLE Plasmodium falciparum 3D7 chromosome 12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 244649)
AUTHORS Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-1998) Stanford DNA Sequencing and Technology
Center, Stanford University, 855 California Avenue, Palo Alto, CA
94304, USA
COMMENT On Aug 20, 1998 this sequence version replaced gi:3334987.
*** WARNING: Phase 1 High Throughput Genome Sequence ***
***
* This sequence is unfinished. It consists of 36 contigs for
* which the order is not known: their order in this record is
* arbitrary. In some cases, the exact lengths of the gaps
* between the contigs are also unknown; these gaps are presented
* as runs of N as a convenience only. When sequencing is complete,
* the sequence data presented in this record will be replaced
* by a single finished sequence with the same accession number.

```

FEATURES
source
1. .244649

1 11189: contig of 11189 bp in length
24361: gap of unknown length
37253: contig of 13172 bp in length
42268: gap of unknown length
51249: contig of 12892 bp in length
58344: gap of unknown length
65377: contig of 5015 bp in length
76265: gap of unknown length
90094: contig of 8981 bp in length
103795: gap of unknown length
113640: contig of 7095 bp in length
122079: gap of unknown length
130477: contig of 8988 bp in length
138589: gap of unknown length
146666: contig of 8440 bp in length
154737: gap of unknown length
162631: contig of 8398 bp in length
170496: gap of unknown length
177936: contig of 8077 bp in length
185316: gap of unknown length
191922: contig of 8071 bp in length
198084: gap of unknown length
203810: contig of 7894 bp in length
208771: gap of unknown length
213163: contig of 7865 bp in length
217459: gap of unknown length
221544: contig of 7440 bp in length
225441: gap of unknown length
228635: contig of 7380 bp in length
231186: gap of unknown length
233697: contig of 6606 bp in length
236048: gap of unknown length
238359: contig of 6162 bp in length
240643: gap of unknown length
242673: contig of 4392 bp in length
244649: gap of unknown length
246649: contig of 4296 bp in length
248649: gap of unknown length
250649: contig of 4085 bp in length
252649: gap of unknown length
254649: contig of 3897 bp in length
256649: gap of unknown length
258649: contig of 3194 bp in length
260649: gap of unknown length
262649: contig of 2551 bp in length
264649: gap of unknown length
266649: contig of 2511 bp in length
268649: gap of unknown length
270649: contig of 2351 bp in length
272649: gap of unknown length
274649: contig of 2311 bp in length
276649: gap of unknown length
278649: contig of 2284 bp in length
280649: gap of unknown length
282649: contig of 2030 bp in length
284649: gap of unknown length
286649: contig of 1976 bp in length.

[illegible]

KEYWORDS pyk10 gene; thioglucosidase; thioglucoside glucohydrolase.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
 Eukaryotes; mitochondrial eukaryotes; Viridiplantae;
 Charophyta/Embryophyta group; Embryophyta: Magnoliophyta;
 Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1744)
AUTHORS Schmidt, K.P., Burrows, P.R., Davies, K.G., Kammerloher, W.,
 Schaeffner, A.R., Buck, F., Cal.D. and Grundler, F.M.W.
TITLE A root specific myrosinase in Arabidopsis responding to cyst
 nematode infection
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1744)
AUTHORS Schmidt, K.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-1995) K.P. Schmidt, Institut fuer
 Phytopathologie, Hermann-Rodewald-Str. 9, D- 24118 Kiel, FRG
COMMENT Related sequences: AT4257 and Z34217.
FEATURES Location/Qualifiers
 source
 1..1744
 /organism="Arabidopsis thaliana"
 /strain="Landsberg erecta"
 /db_xref="taxon:3702"
 /dev_stage="28 days old"
 /tissue_type="root tissue"
 /clone_lib="pCDM8 COS cell expression library"
 /clone="full length in pBsc M13"
 6..77
 /gene="pyk 10"
 6..1580
 /gene="pyk 10"
 /EC_number="3.2.3.1"
 /codon_start=1
 /product="thioglucoside glucohydrolase"
 /db_xref="PID:g984052"
 /translation="MVLQKPLIGLLITVASPNADGVCPSPNKLRSAPPEGF
 LFGTAAYOVEGAINTCRGPALWDYICRRYPERCNDNGDVAVFHRRKEDIQLM
 KNLTDAPRSLIAWPRIFPHGRKEGVSQAGVQFYHDLINXGIPFPYVFEMDT
 QNLDIEYGGFLSRIVKDFREYADVFQYGGVKKWITFNEPWFVLAGYDVGKKA
 PGRCSYVNAKQDGRSGYRAYLVTHNLLISHAEVAYRCKEKGKIGIAUSPAW
 FEALDADSGASIDRALDFILGHLDLTTFTGDPQIMKDIVGHRLPKFTTEOKAL
 KASIDFGLNYISVFNHLEKPPSPRNMQDSLITHEKNAQYALGSKPLTAALN
 VYSGRSLKLIKDKTANPEIMENGVEELGASDSVAVGTADHNRRKLYLORLLS
 MQEAVCDKVNVTGYFYWSLLDNPEWQGYKNRFLYTVDFKNLTRYEKESGYIKD
 FLQGVPRPSALKKDEL"
 6..1580
 /gene="pyk 10"
 polyA_signal 561 a 332 g 380 g 471 t
BASE COUNT 561 a 332 g 380 g 471 t
ORIGIN
 Query Match 5.3%; Score 46; DB 23; Length 1744;
 Best Local Similarity 91.1%; Pred. No. 4,97e-02;
 Matches 51; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 Db 1654 AATAAAGGCGTATTATTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1709
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 810 aataaataaggtattattttataaataaataaataaataaataaataa 865
gene
 polyA_signal 561 a 332 g 380 g 471 t
BASE COUNT 561 a 332 g 380 g 471 t
ORIGIN
 Query Match 5.3%; Score 46; DB 23; Length 1744;
 Best Local Similarity 91.1%; Pred. No. 4,97e-02;
 Matches 51; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 Db 1654 AATAAAGGCGTATTATTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1709
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 810 aataaataaggtattattttataaataaataaataaataaataaataa 865
gene
 polyA_signal 561 a 332 g 380 g 471 t
BASE COUNT 561 a 332 g 380 g 471 t
ORIGIN
 Query Match 5.3%; Score 46; DB 28; Length 1999;
 Best Local Similarity 91.1%; Pred. No. 4,97e-02;
 Matches 51; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 Db 1942 AATAAAGGCGTATTATTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1997
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 810 aataaataaggtattattttataaataaataaataaataaataaataa 865
RESULT 13
LOCUS BTU73585 3795 bp mRNA MAM 15-AUG-1997
DEFINITION Bos taurus Fanconi anemia group C mRNA, complete cds.
ACCESSION U73585
NID 92326995
KEYWORDS Bos taurus.
SOURCE Bos taurus.
ORGANISM Bos taurus
 Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Mammalia; Eutheria; Artiodactyla; Ruminantia; Pecora;
 Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 3795)
AUTHORS Wong, J.C.Y., Alon, N. and Buchwald, M.
TITLE Cloning of the bovine and rat Fanconi anemia group C cDNA
JOURNAL Mamm. Genome 8 (7), 522-525 (1997)
MEDLINE 97341119
REFERENCE 2 (bases 1 to 3795)
AUTHORS Wong, J.C.Y., Buchwald, M. and Alon, N.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1996) Genetics, Hospital for Sick Children, 555
 University Ave., Toronto, Ontario M5G 1X8, Canada
FEATURES Location/Qualifiers
 source
 1..3795
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /tissue_type="Liver"
 727..2430
 /note="similar to human PAC protein, GenBank Accession
 Numbers X66893 and X66894"
 /codon_start=1
 /product="Fanconi anemia group C protein"
 /db_xref="PID:g2326996"
CDS

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1999)
AUTHORS Johnson, C.A., White, D., O'Neill, L.P. and Turner, B.M.
TITLE Direct Submission
JOURNAL Submitted (02-DEC-1996) Department of Anatomy, University of
 Birmingham, Birmingham B15 2TT, U.K.
COMMENT On Sep 2, 1997 this sequence version replaced gi:1737470.
FEATURES Location/Qualifiers
 source
 1..1999
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 1..1999
 /gene="Hdl"
 15..1463
 /gene="Hdl"
 /note="similar to human histone deacetylase HD1 and yeast
 RPD3"
 /codon_start=1
 /product="putative histone deacetylase"
 /db_xref="PID:g2347180"
 /translation="MAQTGKTKKVCYYDGVNYYGCGHPMPXPHRIRMTNHLN
 YGLYKMEIYRPHKANAEPMKYHSDDIKELRSIRPDNMSEYKMQMORFNVGDCPV
 FDLGFEFCQLSTGGSVASAVKLKQOTDIQVNWAGLHAKKSEASGFCYVNDIVLAI
 LELLKYHQRVLYIDIDIHGQGVZAFYTTDRVTVYSPHKYGEYFPGDLDIGAGK
 GKYAVNTPLRDGDDESEIALFKPMVSKVMEMFQSAVVLQCGSDLSGDLGCFNL
 TIKGHAKCVFVKSFNLMNLGGGGYTIIRVARCWTYETAVLDEITPNEIYNDYF
 EYFGDFKLHISPSNMTNINLEYLKIQLRLENLRLMPLRPHAPGVQMAIPEDAIPEE
 SGDEDEPDKRIKISDCKRIACEEFSDESGEGGKKNSSNFKKARVKTDEKE
 KDPEKKEVTEETKEKPEAKVEKVKLA"
 583 a 467 c 480 g 467 t 2 others
BASE COUNT 583 a 467 c 480 g 467 t
ORIGIN
 Query Match 5.3%; Score 46; DB 28; Length 1999;
 Best Local Similarity 91.1%; Pred. No. 4,97e-02;
 Matches 51; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 Db 1942 AATAAAGGCGTATTATTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1997
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 810 aataaataaggtattattttataaataaataaataaataaataaataa 865
RESULT 13
LOCUS BTU73585 3795 bp mRNA MAM 15-AUG-1997
DEFINITION Bos taurus Fanconi anemia group C mRNA, complete cds.
ACCESSION U73585
NID 92326995
KEYWORDS Bos taurus.
SOURCE Bos taurus.
ORGANISM Bos taurus
 Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Mammalia; Eutheria; Artiodactyla; Ruminantia; Pecora;
 Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 3795)
AUTHORS Wong, J.C.Y., Alon, N. and Buchwald, M.
TITLE Cloning of the bovine and rat Fanconi anemia group C cDNA
JOURNAL Mamm. Genome 8 (7), 522-525 (1997)
MEDLINE 97341119
REFERENCE 2 (bases 1 to 3795)
AUTHORS Wong, J.C.Y., Buchwald, M. and Alon, N.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1996) Genetics, Hospital for Sick Children, 555
 University Ave., Toronto, Ontario M5G 1X8, Canada
FEATURES Location/Qualifiers
 source
 1..3795
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /tissue_type="Liver"
 727..2430
 /note="similar to human PAC protein, GenBank Accession
 Numbers X66893 and X66894"
 /codon_start=1
 /product="Fanconi anemia group C protein"
 /db_xref="PID:g2326996"
CDS

polyA_site	3759	BASE COUNT	1023 a	868 c	886 g	1018 t	ORIGIN
Query Match	5.3%	Score 46:	DB 19:	Length 3795;			
Best Local Similarity	87.1%	Pred. No.	4.97e-02:				
Matches	54;	Conservative	0;	Mismatches 8;	Indels 0;	Gaps 0;	
Db	3732	CAATAATAAAACAGGACAGTATTCTAA	AAAAAAAAAAAAAAAAAAAAAAAAAAAAA	AAAAAAAAAAAAAAAAAAAAAAAAAAAAA	AAAAAAAAAAAAAAAAAAAAAAAAAAAAA	AAAAAAAAAAAAA	3791
Qy	804	caataataaataaggtattatttataa	aaaaaaaaaaaaaaaaaaaaaaaaaaaaa	aaaaaaaaaaaaaaaaaaaaaaaaaaaaa	aaaaaaaaaaaaa	aaaaa	863
Db	3792	AA	3793				
Qy	864	aa	865				
RESULT	14	AF012089	10772 bp	DNA	INV	05-AUG-1997	
LOCUS	DEFINITION	Drosophila melanogaster cysteine proteinase-1 (CPI) gene, complete cds, and phenylalanine tRNA synthetase gene, partial cds.					
ACCESSION	AF012089						
NID	g2305220						
KEYWORDS		fruit fly.					
SOURCE	ORGANISM	Drosophila melanogaster					
REFERENCE	AUTHORS	Gray,Y.H., Tanaka,M.M. and Sved,J.A.					
TITLE		P-element-induced recombination in Drosophila melanogaster: hybrid element insertion					
JOURNAL	MEDLINE	Genetics 144 (4), 1601-1610 (1996)					
REFERENCE	AUTHORS	Gray,Y.H.M., Sved,J.A., Preston,C.R. and Engels,W.R.					
TITLE		Structure of the cysteine proteinase (CPI) gene of Drosophila melanogaster and associated mutational effects					
JOURNAL		Unpublished					
REFERENCE	AUTHORS	Gray,Y.H.M., Sved,J.A., Preston,C.R. and Engels,W.R.					
TITLE		Direct Submission					
JOURNAL		Submitted (30-JUN-1997) School of Biological Sciences, University of Sydney, Biology A12, Sydney University, NSW 2006, Australia					
FEATURES	source	Location/Qualifiers					
		1..10772					
		/organism="Drosophila melanogaster"					
	mRNA	/db_xref="taxon:7227"					
		join(872)..1000,2310..2426,6476..6690,6751..7707)					
		/gene="Cpi"					
	gene	/product="cysteine protease"					
		872..7707					
		/gene="CPI"					
	exon	872..1000					
		/gene="CPI"					
		/number=1					
	intron	1001..2309					
		/gene="CPI"					
	exon	2310..2426					
		/gene="CPI"					
		/number=2					

W P S R E H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Tue Dec 15 18:57:53 1998; Maspar time 127.57 Seconds
Tabular output not generated. 922.536 Million cell updates/sec

Title: >US-09-109-864-1
Description: (1-865) from US09109864.seq
Perfect Score: 865
N.A. Sequence: 1 tttgtgagggcgcgcagaga.....aaaaaaaaaaaaaaaaaaaa 865
Comp: aaacactccgcgtctct.....tttttttttttttttt

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 188442 seqs, 68026449 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-gensseq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40

Statistics: Mean 9.274; Variance 8.844; scale 1.049

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	47	5.4	766	3	N40162	Sequence of prepropar 5.38e-07
2	46	5.3	91	9	Q51746	Oligonucleotide probe 1.28e-06
3	45	5.2	3437	21	T31478	Spinach debranching e 3.02e-06
4	44	5.1	1066	8	Q49943	Human anti-HBs light 7.13e-06
5	44	5.1	8920	11	Q82924	Carbamoyl-phosphate-s 7.13e-06
6	43	5.0	340	16	Q99393	Rat allograft inflam 1.67e-05
7	43	5.0	350	16	Q99380	Rat allograft inflam 1.67e-05
8	43	5.0	1078	4	Q25785	Protein synthesis inh 1.67e-05
9	43	5.0	1136	2	N70435	Sequence encoding ins 1.67e-05
10	43	5.0	1364	40	V15228	Heterodimeric human 1 1.67e-05
11	43	5.0	1364	26	T48099	Human interleukin-12 1.67e-05
12	43	5.0	1396	4	Q24371	Natural killer cell s 1.67e-05
13	43	5.0	1558	2	Q10896	Encodes Xenopus Bone 1.67e-05

14	43	5.0	3581	10	Q44391	Sequence of murine OS 1.67e-05
15	42	4.9	259	38	V00423	3' fragment of clone 3.91e-05
16	42	4.9	615	38	V02881	Human HMG1-C aberrant 3.91e-05
17	42	4.9	756	12	Q73729	Rape abscission/dehls 3.91e-05
18	42	4.9	1558	22	T28255	Survival motor neuron 3.91e-05
19	42	4.9	1560	22	T18828	Human survival motor 3.91e-05
20	42	4.9	1582	22	T18831	Human survival motor 3.91e-05
21	42	4.9	1582	22	T28259	Survival motor neuron 3.91e-05
22	42	4.9	1710	8	Q47966	Rape acyl-ACP thioest 3.91e-05
23	42	4.9	1738	10	Q56241	Sequence encoding ant 3.91e-05
24	42	4.9	3073	3	Q14635	Clone associated with 3.91e-05
25	42	4.9	3073	21	T34371	Plasmid PATG29 (ATCC 3.91e-05
26	41	4.7	198	7	Q42784	Ligand-induced gene, 9.09e-05
27	41	4.7	259	38	T91303	Human HI075-1 secrete 9.09e-05
28	41	4.7	688	18	T16856	Integrin subunit beta 9.09e-05
29	41	4.7	867	38	V02135	Human secreted protei 9.09e-05
30	41	4.7	867	39	T88057	Partial cDNA clone en 9.09e-05
31	41	4.7	877	39	V04635	Human cytoplasmic dom 9.09e-05
32	41	4.7	1420	35	T72167	Alzheimer's disease D 9.09e-05
33	41	4.7	1441	38	T90173	Oil seed rape cystein 9.09e-05
34	41	4.7	1474	38	T90174	Oil seed rape cystein 9.09e-05
35	41	4.7	1569	1	N90711	Recombinant activatio 9.09e-05
36	41	4.7	1640	9	Q50575	Asparaginylendopeptid 9.09e-05
37	41	4.7	1910	9	Q50579	Asparaginylendopeptid 9.09e-05
38	41	4.7	2116	39	T99541	Human hyaluronan synt 9.09e-05
39	41	4.7	2389	30	T61590	Human c-IAP1. 9.09e-05
40	41	4.7	3504	1	Q05304	Clone lmd2 encoding i 9.09e-05
41	41	4.7	3850	1	N81634	Sequence of new plasm 9.09e-05
42	41	4.7	3851	33	T79857	Yellowtail tuna DNA e 9.09e-05
43	41	4.7	3933	16	T03885	Human mucosal lymphoc 9.09e-05
44	41	4.7	5108	9	Q53403	Sequence encoding ail 9.09e-05
45	41	4.7	8239	29	T61981	Human ataxia and rad 9.09e-05

ALIGNMENTS

RESULT 1
ID N40162 standard; cDNA; 766 BP.
AC N40162;
DT 16-FEB-1992 (first entry)
DE Sequence of preproparathyroid cDNA.
KW Proteolytic cleavage; trypsin; protprotein; precursor; proteolytic;
KW processing; ss.
OS Bos taurus.
FH Key Location/Qualifiers
FT cds 111..460
ET /*tag= a
FN W08401173-A.
PD 29-MAR-1984.
PF 09-SEP-1983; U01361.
PR 15-SEP-1982; US-418537.
PR 22-DEC-1982; US-452339.
PA (IMMU-) IMMUNO NUCLEAR CORP.
PI Mahoney WC;
DR WPI; 84-088522/14.
DR P-PSDB; P40209.
PT Prodn. of protein in yeast transformed to express precursor - by
PT use of enzyme(s) naturally present in yeast
PS Example; Fig 2; 28pp; English.
CC The inventors claim plasmid pYEM-1 and yeast transformed by a plasmid
CC pYEM-1. The yeast is pref. Saccharomyces cerevisiae or S. pombe.
CC Yeast strain X1069-2b transformed by plasmid pYEM-1 is deposited as
CC NRRL Y-15153. pYEM-1 is constructed by the ligation of a DNA
CC fragment (obtd. in stages from a cDNA sequence coding for bovine
CC preproparathyroid hormone) into the Hind III site of modified
CC plasmid vep-13.
SQ Sequence 766 BP; 262 A; 142 C; 158 G; 204 T;

Query Match 5.4%; Score 47; DB 3; Length 766;
Best Local Similarity 91.2%; Pred. No. 5.38e-07;
Matches 52; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Db 664 aaataaatgaagttaagtattataaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 720

```

Query Match      5.1%; Score 44; DB 8; Length 1066;
Best Local Similarity 86.7%; Pred. No. 7.13e-06;
Matches 52; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db      912  aatgaataataaagtgaaattcttgcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 971
      ||| ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      806  aataataataaaggtattattttaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 865

RESULT# 5
ID      O62924 standard; cDNA; 8920 BP.
AC      O62924;
DT      O6-DEC-1994 (first entry)
DE      Carbamoyl-phosphate-synthetase II.
KW      Carbamoyl-phosphate-synthetase II; CPSII; pscPSII gene;
      malaria; ss.
OS      Plasmodium falciparum.
Key      Location/Qualifiers
FH      1236..8401
cds      /*tag= a
      /EC_number= 6.3.5.5

WO9412643-A.
09-JUN-1994.
02-DEC-1993; AU0617.
03-DEC-1992; AU-006206.
16-DEC-1992; AU-006380.
(LUNX) UNISEARCH LTD.
PI      Flores MV, Osullivan WJ, Stewart TS;
DR      WPI; 94-200271/24.
PT      Nucleic acid encoding carbamoyl phosphate synthetase II -

```

```

08-FEB-1996      (first entry)
DT Rat allograft inflammatory factor 2 (AIF-2) cDNA fragment.
DE AIF-2; allograft inflammatory factor 2; transplant rejection;
KW inhibitor; immunogenic; detection; diagnosis; ds.
QS Rattus sp.
OS W09517506-AI.
PN 29-JUN-1995.
PD PF 21-DEC-1994; U14724.
PR PA 21-DEC-1993; US-171385.
(PHARD ) HARVARD COLLEGE.
PI Russell ME   Utans U;
DR WPI; 95-240668/31.
PT DNA encoding allograft rejection factors and immunogenic fragments - useful for identifying transplant rejection inhibitors
PS Claim 11: Page 92; 138pp; English.
CC Q99372-Q99393 are rat allograft inflammatory factor 2 (AIF-2) cDNA fragments. The AIF-2 gene is a differentially expressed allograft gene which is expressed in allograft tissue during transplant rejection. Identification of the rat AIF-2 product (#80521) or transcript indicates that allograft rejection is taking place. The rat AIF-2 gene, fragments of this gene and/or the gene product are therefore useful in the diagnosis of transplant rejection.
CC The diagnostic methods used allow rejection (vascular inflammation) to be detected at an early stage and require only a small amount of biopsy material.
CC (sequence given in sequence ID listing in the specification gives nucleotide 111 as C)
CC Sequence 350 BP;    121 A;    37 C;    63 G;    129 T;
SQ
Query Match          5.0%; Score 43; DB 16; Length 350;
Best Local Similarity 78.7% ; Pred. No. 1.67e-05;
Matches 59; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Db       2 ttttttttttttttttttttttttttttttttttttttttttcatatcagtagtttatct 61
Cp       865 ttutttttttttttttttttttttttttttttttttaataaacacttttttttttt 806

Db        62 tagccacaatatgaagt 76
Cp       805 tgaccacaaataaat 791

RESULT 8
ID     ID Q25785 standard; DNA; 1078 BP.
AC     AC Q25785;
DT     DT 10-DSC-1992 (first entry)
DE     DE Protein synthesis inhibiting gene.
KW     KW Plant pathogens; pathogen resistance; medicine; ss.
Key    Key Hordeum vulgare.
FH     FH Location/Qualifiers
FT     FT cds         43..888
FT     FT /*tag= a
FT     FT 5'utr       1..42
FT     FT /*tag= b
FT     FT 3'utr       886..1078
FT     FT /*tag= c
FT     FT poly_a_signal 930..935
FT     FT /*tag= d
FT     FT /note= "possible polya signal"
FT     FT poly_a_signal 963...976
FT     FT /*tag= e
FT     FT /note= "possible polya signal"
FT     FT poly_a_signal 1002..1011
FT     FT /*tag= f
FT     FT /note= "possible polya signal"
FT     FT poly_site    1032..1078
FT     FT /*tag= g
DE4040954-A.
PN     PN 25-JUN-1992.
PD     PD 20-DEC-1990; DE-040954.
PR     PR (PLAC ) MAX PLANCQ GES FORDERUNG WISSE.
PA     PA Jach g., Iogemann J., Mundy J., Schell J.;
```



```

DR WPI: 92-217825/27.
DR P-PSDB; R25048.
PT Pathogen resistant plants prodm. using protein synthesis inhibiting
PT gene - or its fusion products, inserted into the genome, also
PT pharmaceutical use of derived synthesis inhibiting protein
PS Disclosure; Fig 3; 23pp; German.
CC The sequence is that of a gene encoding a protein which effectively
CC blocks protein synthesis by plant pathogens, e.g. Trichoderma reesi
CC and Fusarium sporotrichoides. It can be fused to active promoters
CC such as wun-1 and the fusion introduced into plant genetic material
CC to impart pathogen resistance to both mono- and di-cotyledonous
CC plants. The gene was isolated from mature barley seeds.
SQ Sequence 1078 BP; 276 A; 310 C; 321 G; 171 T;

Query Match 5.0%; Score 43; DB 4; Length 1078;
Best Local Similarity 85.2%; Pred. No. 1.67e-05;
Matches 52; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Db 1005 aaataaataataaataatgcagtttaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1064
QY 805 aaataaataataaagttattattataaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 864
Db 1065 a
QY 865 a 865

RESULT 9
ID N70435 standard; cDNA; 1136 BP.
AC N70435;
DE 05-APR-1991 (first entry)
DE Sequence encoding insulin-like growth factor 1B (IGF-1B).
KW Growth promoter; lactation enhancer; cell proliferation; ss.
OS Homo sapiens.
PN EP-229750-A.
PD 22-JUL-1987.
PF 06-JAN-1987; 870001.
PR 07-JAN-1986; US-816662.
PR 20-NOV-1986; US-929671.
PA (UNIW ) UNIV OF WASHINGTON.
PI Krivi GG. Rotwein PS;.
DR WPI: 87-200203/29.
PT New pre-pro-insulin-like growth factor-1 protein - obtd. by
PT recombinant DNA procedures for use as growth promoters for
PT enhancing lactation, for stimulating cell proliferation etc.
PS Example; Fig 5; 59pp; English.
CC A 42 base oligonucleotide corresponding to the DNA sequence encoding
CC amino acids 10 to 23 of mature human IGF-1 was synthesized (N70437).
CC The radiolabeled 42 mer was then employed to screen for IGF-1
CC containing DNA sequences in a human liver cDNA library. Insulin-
CC like growth factors-1A and -1B cDNAs were isolated from a human
CC library by using lambdaclat 11 (N70435, N70436). The human IGF-1
CC genomic gene was isolated and mapped. It encodes at least two
CC preproinsulin-like growth factor-1 proteins. An essentially pure
CC proproinsulin-like growth factor-1 protein comprising the sequence
CC of amino acids shown in Figure six is claimed (P70277).
SQ Sequence 1136 BP; 412 A; 230 C; 268 G; 226 T;

Query Match 5.0%; Score 43; DB 2; Length 1136;
Best Local Similarity 87.7%; Pred. No. 1.67e-05;
Matches 50; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 1073 aaataaataaagattataaataaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1129
QY 809 aaataaataaagttattattataaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 865

RESULT 10
ID V15228 standard; cDNA; 1364 BP.
AC V15228;
DE 27-MAY-1998 (first entry)
DE Heterodimeric human interleukin 12 subunit 2 encoding cDNA.
KW Heterodimeric; human; interleukin 12; IL-12; vaccine; antigen;

```

```

KW pathogen; immune response; microorganism; cancer; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
CDS 101..862
FT /*tag= a
FT /product= "heterodimeric human interleukin 12 subunit"
PD US5723127-A.
PN 03-MAR-1998.
PF 25-MAR-1996; 621493.
PR 17-JUN-1994; US-265087.
PR 18-APR-1994; US-229282.
PR 25-MAR-1996; US-621493.
PA (TYPE-) UNIV PENNSYLVANIA.
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
PI Scott P, Trinchieri G;
DR WPI: 98-205996/18.
DR P-PSDB; W44325.
DR Enhancing immune response to vaccine - comprises co-administering
DR interleukin-12 with antigenic composition
PT Interleukin-12 with antigenic composition
PS Disclosure; Column 23-26; 23pp; English.
CC The present sequence encodes a heterodimeric human interleukin 12
CC (IL-12) subunit (putatively the 30 kD subunit) shown in the present
CC specification. The present invention describes a method for enhancing
CC the immune response elicited by an antigenic composition. The method
CC comprises administering IL-12 simultaneously or sequentially with the
CC antigenic composition. The IL-12 can be used in therapeutic
CC compositions, e.g. as an adjuvant in vaccines against pathogenic
CC bacteria or viruses, especially in an amount of 0.1 mu g to 0.5 mg per
CC 25 mu g of antigen, or in cancer vaccines.
SQ Sequence 1364 BP; 446 A; 282 C; 275 G; 361 T;

Query Match 5.0%; Score 43; DB 40; Length 1364;
Best Local Similarity 90.6%; Pred. No. 1.67e-05;
Matches 48; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 1299 aaataaagattgattataaataaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1351
QY 813 aaataaggtattattattataaataaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 865

RESULT 11
ID T48099 standard; cDNA; 1364 BP.
AC T48099;
DE 21-MAR-1997 (first entry)
DE Human interleukin-12 30 kD subunit gene.
KW Human; interleukin 12; IL-12; heterodimeric complex; immunogenic; ds;
KW Leishmania; antigen; leishmaniasis; cell-mediated immunity; adjuvant;
KW TH1 helper cell; TH2 helper cell; alum; cytokine; bacterium; cancer;
KW vaccine; cell surface; membrane; glycoprotein; antigen presenting cell.
OS Homo sapiens.
FH Key Location/Qualifiers
CDS 101..862
FT /*tag= a
FT /product= interleukin-12 30 kD subunit
PD US571515-A.
PN 05-NOV-1996.
PF 18-APR-1994; 229282.
PR 18-APR-1994; US-229282.
PR 17-JUN-1994; US-265087.
PA (TYPE-) UNIV PENNSYLVANIA.
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
PI Scott P, Trinchieri G;
DR WPI: 96-505347/50.
DR P-PSDB; W07399.
DR Immunogenic compsn. to improve cell mediated immune response -
PT contains soluble leishmania antigen and interleukin-12 as adjuvant
PS Disclosure; Column 23-26; 22pp; English.
CC This is the nucleotide sequence encoding the human interleukin (IL)-12
CC 30 kD subunit. IL-12 is a heterodimeric complex composed of the 30 kD
CC subunit and a 40 kD subunit (W07398). The complex is used in a novel
CC immunogenic composition comprising a soluble leishmania antigen with
CC IL-12, for protection against leishmaniasis. The addition of IL-12
CC improves cell-mediated immunity by inducing TH1 helper cells (as opposed

```

CC to TH2 cells which are induced by alum adjuvant) and does not cause
 CC uncontrolled release of other cytokines (in contrast to bacterial
 CC adjuvants). IL-12 can also be used as a cancer vaccine by association
 CC with the protein B7, a soluble, cell-surface (membrane)-bound
 CC glycoprotein which is expressed in antigen presenting cells.
 SQ Sequence 1364 BP; 446 A; 282 C; 275 G; 361 T;

Query Match 5.0%; Score 43; DB 26; Length 1364;
 Best Local Similarity 90.6%; Pred. No. 1.67e-05;
 Matches 48; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 1299 aaataaagtattgattataaaataaaataaaataaaataaaataaaataaa 1351
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 813 aaataagttattttataaaataaaataaaataaaataaaataaaataaa 865

RESULT 12

ID Q24371 standard; DNA; 1396 BP.

AC Q24371;

DE 16-SEP-1992 (first entry)

DT Natural killer cell stimulatory factor 30kd subunit.

KW NKSF; cytokine; gamma interferon; IFN; peripheral blood lymphocytes;

IL-2; GM-CSF; granulocyte macrophage colony stimulating factor; PBL;

QY Homo sapiens.

FT Key Location/Qualifiers

FT misc_feature 14..1364

FT /tag= a

FT /label= p35nksf14-1-1

FT /note= "long clone containing 30kd NKSF subunit"

FT misc_feature 146..1333

FT /tag= b

FT /label= p35nksf9-1-1

FT /note= "short clone containing 30kd NKSF subunit"

FT cds 113..873

FT /tag= c

FT /product= pro NKFS 30kd subunit

PN W9205256-A.

PD 02-APR-1992.

PF 04-SEP-1991; U06332.

PR 18-SEP-1990; US-584941.

PA (GENE-) GENETICS INST INC.

PI (WIST-) WISTAR INST.

PI Clark S C, Hewick R, Kobayashi M, Perussia B, Trinchieri G;

PI Wong G G;

DR WPI: 92-132124/16.

DR P-PSDB: R23729.

PT New natural killer cell stimulatory factor - useful as an

PT immunostimulant for inducing gamma-interferon and GM-CSF prodn.

PT in human peripheral blood lymphocytes

PS Claim 12: Page 23; 79pp; English.

CC NKSF is capable of inducing the production of gamma interferon in

CC human peripheral blood lymphocytes. It has subunits of 40 (R22769)

CC and 30-35 kD (R23729). It is claimed that NKSF is useful in the

CC treatment of bacterial and viral infections e.g. AIDS.

CC Degenerate probes were designed from tryptic fragments of NKSF

CC and used to screen a cDNA library made from poly A RNA from pBbu

CC induced 8866 cells to identify a cDNA encoding the small subunit

CC of NKSF.

CC See also R22769, Q23586, R23729.

SQ Sequence 1396 BP; 451 A; 294 C; 285 G; 366 T;

Query Match

Best Local Similarity 90.6%; Pred. No. 1.67e-05;

Matches 48; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 1312 aaataaagtattgattataaaataaaataaaataaaataaaataaa 1364
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 813 aaataagttattttataaaataaaataaaataaaataaaataaa 865

RESULT 13

ID Q10896 standard; DNA; 1558 BP.

AC Q10896;

DT 13-MAY-1991 (first entry)

DE Encodes Xenopus Bone Morphogenetic Factor BMP-2B.

KW BMF; osteoporosis; fracture; cartilage; ss.

OS Xenopus laevis.

FT Key Location/Qualifiers

FT cds 105..1309

FT /tag= a

FT /product= BMP-2B

PN EP-416578-A.

PD 13-MAR-1991.

PF 05-SEP-1990; 117079.

PR 06-SEP-1989; JP-229250.

PR 20-JUL-1990; JP-190774.

PA (TAKE) TAKEDA CHEMICAL IND KK.

PA (SCIT-) SCITECH RESEARCH CO.

PI Murakami K, Ueno N, Kato Y;

DR WPI: 91-075112/11.

DR P-PSDB; R10996.

PT Xenopus laevis bone morphogenetic protein and DNA encoding it -

PT used in therapy of fracture or osteoporosis

PS Claim 5; Fig 2; 28pp; English.

CC A Xenopus laevis unfertilised egg cDNA library in lambda gt10 was

CC screened with a PstI-HindIII fragment of X.laevis Xar14 chromosomal

CC DNA. Three clones were isolated, including clone Xbr23 which was

CC found to encode a protein having homology with X.laevis BMP-2B. They

CC were subcloned in pUC19 and used to transform competent E.coli HB101

CC cells. Transformant E.coli HB101/pXbr23 coding for the BMP-2B was

CC sequenced.

CC See also Q10890-5 and Q10897.

SQ Sequence 1558 BP; 504 A; 327 C; 348 G; 379 T;

Query Match

Best Local Similarity 85.2%; Pred. No. 1.67e-05;

Matches 52; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Db 1467 aaaaaataaataagtcattatttttaaacataaaaaataaaaaataaaaa 1526
 ||| ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| |||||

QY 805 aaataaataaataaggtatttttttttttttttttttttttttttttt 864

Db 1527 a 1527

QY 865 a 865

RESULT 14

ID Q44391 standard; cDNA to mRNA; 3581 BP.

AC Q44391;

DT 14-SEP-1994 (first entry)

DE Sequence of murine OSF-4 cDNA.

KW OSF-4; cadherin; growth factor; osteogenesis; osteoblast; therapy;

OS diagnosis; ss.

OS Mus musculus.

FT Key Location/Qualifiers

FT cds 284..2671

FT /tag= a

PN EP-585801-A.

PD 09-MAR-1994.

PF 23-AUG-1993; 113602.

PR 28-AUG-1992; JP-230028.

PA (FARR) HOECHST JAPAN LTD.

PI Amann E, Kawai S, Okazaki M, Takeshita S, Tsujimura A;

DR WPI: 94-076152/10.

DR P-PSDB; R49730.

PT New bone related, cadherin-like OSF-4 proteins - for treatment

PT and diagnosis of bone metabolic disease, and nucleic acid

PT encoding them

PS Claim 3; Page 13-17; 34pp; English.

CC cDNA libraries were constructed from the mouse osteoblastic cell

CC line MC3T3-E1 and from mouse liver tissue, amplified by PCR, and

CC then as much common DNA as possible removed by hybridisation between

CC the 2 libraries. Residual E1-specific DNA was amplified, inserted

CC into lambda gt10 and screened for plaque hybridisation. A minibank of

Search completed: Tue Dec 15 19:01:29 1998
Job tlme : 216 secs.

Query Match 4.9%; Score 42; DB 38; Length 259;
Best Local Similarity 78.7%; Pred. NO. 3.9le-05;
Matches 59; Conservative 0; Mismatches 15; Indels 1; Caps 1;

Db	159	attnanttttgnataataaa-aaagctgttgtttgtccaaaaaataaaaaaaa	217
QY	791	attatttgggtcgaataaaataaaggtattattataaaaaaataaaaaaaa	850
Db	218	aaaaaataaaaaaaa	232
QY	851	aaaaaataaaaaaaa	865

Result No.	Query	Score	Match	Length	DB	ID	Description	Pred. No.
c 1	44	5.1	7218	1	US-08-232-	Sequence 14, Applicati	9.26e-09	
c 2	43	5.0	340	1	US-08-171-	Sequence 27, Applicati	2.65e-08	
c 3	43	5.0	350	1	US-08-171-	Sequence 14, Applicati	2.65e-08	
c 4	43	5.0	1364	1	US-08-265-	Sequence 3, Applicatio	2.65e-08	
c 5	43	5.0	1558	1	US-08-455-	Sequence 7, Applicatio	2.65e-08	
c 6	43	5.0	7218	1	US-08-232-	Sequence 14, Applicati	2.65e-08	
c 7	42	4.9	3072	1	US-07-688-	Sequence 31, Applicati	7.52e-08	
c 8	42	4.9	3073	2	PCT-US91-0	Sequence 30, Applicati	7.52e-08	
c 9	41	4.7	198	2	PCT-US92-1	Sequence 16, Applicati	7.13e-07	
c 10	41	4.7	688	3	5498694-3	Patent No. 5498694	2.13e-07	
c 11	41	4.7	1049	2	US-08-361-	Sequence 4, Applicatio	2.13e-07	
c 12	41	4.7	2589	2	PCT-US96-1	Sequence 1, Applicatio	2.13e-07	
c 13	41	4.7	3933	1	US-08-199-	Sequence 1, Applicatio	2.13e-07	
c 14	41	4.7	3933	2	PCT-US95-0	Sequence 1, Applicatio	2.13e-07	
c 15	41	4.7	5108	1	PCT-US97-642-	Sequence 1, Applicatio	2.13e-07	
c 16	40	4.6	347	1	US-08-104-	Sequence 2, Applicatio	5.98e-07	
c 17	40	4.6	1283	1	US-08-174-	Sequence 19, Applicati	5.98e-07	
c 18	40	4.6	2158	1	US-08-261-	Sequence 1, Applicatio	5.98e-07	
c 19	40	4.6	2158	1	US-07-602-	Sequence 1, Applicatio	5.98e-07	
c 20	40	4.6	2550	3	5358287-23	Patent No. 5238287	5.98e-07	

```
CC LENGTH: 7218 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC IMMEDIATE SOURCE:
CC CLONE: PTZgpt-F15
SQ SEQUENCE 7218 BP; 1944 A; 1491 C; 1486 G; 1929 T; 368 OTHER.

Query Match 5.1%; Score 44; DB 1; Length 7218;
Best Local Similarity 0.7%; Pred. No. 9.26e-09;
Matches 2; Conservative 165; Mismatches 123; Indels 0; Gaps 0;

Db 1147 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1206
Cc : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cc 865 ttttttttttttttttttttttttttttttttttttttttttttttttttttttt 806
Db 1207 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1366
Cc : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cc 805 tgaccacaaataataataaataaataaataaataaataaataaataaataaataa 746
Db 1267 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1326
Cc : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cc 745 tggagatattttagtctaacactgcctctgtgagtgagtcagtcagccagccctgaag 686
Db 1327 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1386
Cc : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cc 685 cagatggagccatccccaggtcctggtggaggaaggtgcagagtcgaggaaccttgg 626
Db 1387 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1436
Cc : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cc 625 cactctgtgcttccctcagggcccaagctcctgtagactcagtcctcgt 576

RESULT 2
ID US-08-171-385-27 STANDARD; DNA; UNC; 340 BP.
AC xxxxxx
DT
DE Sequence 27, Application US/08171385
CC Sequence 27, Application US/08171385
CC Patent No. 5527884
CC GENERAL INFORMATION:
CC APPLICANT: Mary E. Russell
CC APPLICANT: Ulrike Utans
CC TITLE OF INVENTION: Mediators of Chronic Allograft
CC TITLE OF INVENTION: Rejection
CC NUMBER OF SEQUENCES: 32
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Fish & Richardson
CC STREET: 225 Franklin Street
CC CITY: Boston
CC STATE: Massachusetts
CC COUNTRY: U.S.A.
CC ZIP: 02110-2804
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
CC COMPUTER: IBM PS/2 Model 502 or 55SX
CC OPERATING SYSTEM: MS-DOS (Version 5.0)
CC SOFTWARE: WordPerfect (Version 5.1)
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/171,385
CC FILING DATE:
CC CLASSIFICATION: 514
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Fraser, Janis K.
CC REGISTRATION NUMBER: 34,819
CC REFERENCE/DOCKET NUMBER: 05433/006001
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 542-5070
CC TELEFAX: (617) 542-8906
CC TELEX: 200154
CC INFORMATION FOR SEQ ID NO: 27:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 340
CC TYPE: nucleic acid
CC STRANDEDNESS: double
CC TOPOLOGY: linear
SQ SEQUENCE 340 BP; 114 A; 38 C; 64 G; 124 T; 0 OTHER.

Query Match 5.0%; Score 43; DB 1; Length 340;
Best Local Similarity 78.7%; Pred. No. 2.65e-08;
Matches 59; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Db 4 TTTTtttttttttttttttttttttttttttttttttttttttttttttttttttttt 63
Cc 865 ttttttttttttttttttttttttttttttttttttttttttttttttttttttt 806
Db 64 TAGCCATAATGAAT 78
Cc 805 tgaccacaaataaataa 791

RESULT 3
ID US-08-171-385-14 STANDARD; DNA; UNC; 350 BP.
AC xxxxxx
DT
DE Sequence 14, Application US/08171385
CC Sequence 14, Application US/08171385
CC Patent No. 5527884
CC GENERAL INFORMATION:
CC APPLICANT: Mary E. Russell
CC APPLICANT: Ulrike Utans
CC TITLE OF INVENTION: Mediators of Chronic Allograft
CC TITLE OF INVENTION: Rejection
CC NUMBER OF SEQUENCES: 32
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Fish & Richardson
CC STREET: 225 Franklin Street
CC CITY: Boston
CC STATE: Massachusetts
CC COUNTRY: U.S.A.
CC ZIP: 02110-2804
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
CC COMPUTER: IBM PS/2 Model 502 or 55SX
CC OPERATING SYSTEM: MS-DOS (Version 5.0)
CC SOFTWARE: WordPerfect (Version 5.1)
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/171,385
CC FILING DATE:
CC CLASSIFICATION: 514
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Fraser, Janis K.
CC REGISTRATION NUMBER: 34,819
CC REFERENCE/DOCKET NUMBER: 05433/006001
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 542-5070
CC TELEFAX: (617) 542-8906
CC TELEX: 200154
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 350
CC TYPE: nucleic acid
CC STRANDEDNESS: double
CC TOPOLOGY: linear
SQ SEQUENCE 350 BP; 121 A; 37 C; 63 G; 129 T; 0 OTHER.

Query Match 5.0%; Score 43; DB 1; Length 350;
Best Local Similarity 78.7%; Pred. No. 2.65e-08;
Matches 59; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
```

```

Db      2 TTTTTCCTTTTTTTTTTTTTTTTTTTTTTTTTTTACACACAGTATTTATT 61
Cc      865 tttttttttttttttttttttttttttttttt | ||| | | |||||
Db      62 TAGCCATAAATGAAT 76
Cc      805 tgaccacaataaat 791

RESULT          4
ID US-08-265-087-3 STANDARD; DNA; UNC; 1364 BP.
AC xxxxxx
DT

Sequence 3, Application US/08265087
Cc Sequence 3, Application US/08265087
Cc Patent No. 5571515
Cc GENERAL INFORMATION:
Cc APPLICANT: Scott, Phillip
Cc APPLICANT: Trinchier, Giorgio
Cc TITLE OF INVENTION: Compositions and Methods for Use of
Cc TITLE OF INVENTION: IL-12 as an Adjuvant
Cc NUMBER OF SEQUENCES: 4
Cc CORRESPONDENCE ADDRESS:
Cc ADDRESSEE: Howson and Howson
Cc STREET: Spring House Corporate Center, PO Box 457
Cc CITY: Spring House
Cc STATE: Pennsylvania
Cc COUNTRY: USA
Cc ZIP: 19477
Cc COMPUTER READABLE FORM:
Cc MEDIUM TYPE: Floppy disk
Cc COMPUTER: IBM PC compatible
Cc OPERATING SYSTEM: PC-DOS/MS-DOS
Cc SOFTWARE: Patent In Release #1.0, Version #1.25
Cc CURRENT APPLICATION DATA:
Cc APPLICATION NUMBER: US/08/265,087
Cc FILING DATE:
Cc CLASSIFICATION: 424
Cc PRIOR APPLICATION DATA:
Cc APPLICATION NUMBER: US 08/229,282
Cc FILING DATE: 18-APR-1994
Cc ATTORNEY/AGENT INFORMATION:
Cc NAME: Bak, Mary E.
Cc REGISTRATION NUMBER: 31,215
Cc REFERENCE/DOCKET NUMBER: WST51AUSA
Cc TELECOMMUNICATION INFORMATION:
Cc TELEPHONE: 215-540-9206
Cc TELEFAX: 215-540-5818
Cc INFORMATION FOR SEQ ID NO: 3:
Cc SEQUENCE CHARACTERISTICS:
Cc LENGTH: 1364 base pairs
Cc TYPE: nucleic acid
Cc STRANDEDNESS: double
Cc TOPOLOGY: unknown
Cc MOLECULE TYPE: CDNA
Cc FEATURE:
Cc NAME/KEY: CDS
Cc LOCATION: 101..859
Cc SO SEQUENCE 1364 BP; 446 A; 282 C; 275 G; 361 T; 0 OTHER.

Query Match          5.0%; Score 43; DB 1; Length 1364;
Best Local Similarity 90.6%; Pred.No. 2.65e-08;
Matches 48; Conservative 0; Mismatches 5; Indels 0; Gaps

Db 1299 AAATAAAAGTATGCAATTAAAAAAAAAAAAAAAAAAAAA 1351
Cc      | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Qy 813 aaataagggtattttttaaaaaaaaaaaaaaaaaaaaaaa 865

RESULT          5
ID US-08-455-550-7 STANDARD; DNA; UNC; 1558 BP.
AC xxxxxx
DT

```



```
CC STREET: Two First National Plaza, 20 South Clark
CC STREET: Street
CC CITY: Chicago
CC STATE: Illinois
CC COUNTRY: USA
CC ZIP: 60603
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US91/02714
CC FILING DATE: 19910419
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/511,715
CC FILING DATE: 20-APR-1990
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Borun, Michael F.
CC REGISTRATION NUMBER: 25447
CC REFERENCE/DOCKET NUMBER: 27805/30197
CC TELEPHONE: (312) 346-5750
CC TELEFAX: (312) 984-9740
CC TELEX: 25-3856
CC INFORMATION FOR SEQ ID NO: 30:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 3073 base pairs
CC TYPE: NUCLEIC ACID
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: cDNA
CC NAME/KEY: CDS
CC LOCATION: 3..1111
CC SEQUENCE 3073 BP; 857 A; 684 C; 635 G; 857 T; 0 OTHER.
SQ
Query Match 4.9%; Score 42; DB 2; Length 3073;
Best Local Similarity 88.9%; Pred. No. 7.52e-08;
Matches 48; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 1505 TAAATAATGCTTTGTTGTTAAAAA...AAAAAAAAAAAAAAAAAAAAA 1558
QY 812 taaataagggtattttatttaaaaaa...aaaaaaaaaaaaaaaaaaaaa 865

RESULT 9
ID PCT-US92-10087-16 STANDARD; DNA; UNC; 198 BP.
AC xxxxxx
DE Sequence 16, Application PC/TUS9210087
CC GENERAL INFORMATION:
CC APPLICANT: Smith, Kendall A.
CC TITLE OF INVENTION: IL-2-Stimulated Gene
CC NUMBER OF SEQUENCES: 20
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Lahive & Cockfield
CC STREET: 60 State Street
CC CITY: Boston
CC STATE: Massachusetts
CC COUNTRY: U.S.A.
CC ZIP: 02109
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US92/10087
CC FILING DATE: 19921118

STREET: Two First National Plaza, 20 South Clark
STREET: Street
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/02714
FILING DATE: 19910419
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Borun, Michael F.
REGISTRATION NUMBER: 25447
REFERENCE/DOCKET NUMBER: 27805/30197
TELEPHONE: (312) 346-5750
TELEFAX: (312) 984-9740
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 3073 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
NAME/KEY: CDS
LOCATION: 3..1111
SEQUENCE 3073 BP; 857 A; 684 C; 635 G; 857 T; 0 OTHER.
SQ
Query Match 4.9%; Score 42; DB 2; Length 3073;
Best Local Similarity 88.9%; Pred. No. 7.52e-08;
Matches 48; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 1505 TAAATAATGCTTTGTTGTTAAAAA...AAAAAAAAAAAAAAAAAAAAA 1558
QY 812 taaataagggtattttatttaaaaaa...aaaaaaaaaaaaaaaaaaaaa 865

RESULT 9
ID PCT-US92-10087-16 STANDARD; DNA; UNC; 198 BP.
AC xxxxxx
DE Sequence 16, Application PC/TUS9210087
CC GENERAL INFORMATION:
CC APPLICANT: Smith, Kendall A.
CC TITLE OF INVENTION: IL-2-Stimulated Gene
CC NUMBER OF SEQUENCES: 20
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Lahive & Cockfield
CC STREET: 60 State Street
CC CITY: Boston
CC STATE: Massachusetts
CC COUNTRY: U.S.A.
CC ZIP: 02109
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US92/10087
CC FILING DATE: 19921118

STREET: Two First National Plaza, 20 South Clark
STREET: Street
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/02714
FILING DATE: 19910419
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Borun, Michael F.
REGISTRATION NUMBER: 25447
REFERENCE/DOCKET NUMBER: 27805/30197
TELEPHONE: (312) 346-5750
TELEFAX: (312) 984-9740
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 3073 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
NAME/KEY: CDS
LOCATION: 3..1111
SEQUENCE 3073 BP; 857 A; 684 C; 635 G; 857 T; 0 OTHER.
SQ
Query Match 4.9%; Score 41; DB 2; Length 198;
Best Local Similarity 83.6%; Pred. No. 2.13e-07;
Matches 51; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Db 128 AAATATATTAAAGGAAATGTTTCAGAAAAA...AAAAAAAAAAAAAAAAA 187
QY 805 aaataaataaagggtattttatttaaaaaa...aaaaaaaaaaaaaaaaa 864

Db 188 A 188
QY 865 a 865

RESULT 10
ID 5498694-3 STANDARD; DNA; UNC; 745 BP.
AC xxxxxx
DE 01-JAN-1900
DE Patent No. 5498694.
CC Patent No. 5498694.
CC APPLICANT: RUOSLAHTI, ERKKI I.
CC TITLE OF INVENTION: PEPTIDES OF THE CYTOPLASMIC DOMAIN OF INTEGRIN
CC NUMBER OF SEQUENCES: 12
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/240,967
CC FILING DATE: 10-MAY-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 973,547
CC FILING DATE: 09-NOV-1992
CC APPLICATION NUMBER: 357,824
CC FILING DATE: 25-MAY-1989
CC SEQ ID NO: 3:
CC LENGTH: 688
CC Sequence 745 BP; 234 A; 132 C; 157 G; 165 T; 57 other;
SQ
Query Match 4.7%; Score 41; DB 3; Length 688;
Best Local Similarity 83.6%; Pred. No. 2.13e-07;
Matches 51; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Db 599 AAAATAAATAAAGGCTTACTTCC...AAAAAAAAAAAAAAAAAAAAA 658
QY 805 aaataaataaagggtattttatttaaaaaa...aaaaaaaaaaaaaaaaa 864

Db 659 A 659
QY 865 a 659
```

Qy 865 a 865

RESULT 11
ID US-08-361-467B-4 STANDARD; DNA; UNC; 1046 BP.
AC XXXXXX
DT
DE Sequence 4, Application US/08361467B
CC Sequence 4, Application US/08361467B
CC Patent No. 5633441
CC GENERAL INFORMATION:
CC APPLICANT: De Greef, Willy
CC APPLICANT: Van Emmelo, John
CC APPLICANT: De Oliveria, Dulce E.
CC APPLICANT: De Souza, Maria-Helena
CC APPLICANT: Van Montagu, Marc
CC TITLE OF INVENTION: PLANTS WITH MODIFIED FLOWERS, SEEDS OR
CC TITLE OF INVENTION: EMBRYOS
CC NUMBER OF SEQUENCES: 13
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
CC STREET: P.O. Box 1404
CC CITY: Alexandria
CC STATE: Virginia
CC COUNTRY: United States
CC ZIP: 22313-1404
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/361.467B
CC FILING DATE: 22-DEC-1994
CC CLASSIFICATION: 800
CC PRIOR APPLICATION DATA:
CC PRIOR APPLICATION NUMBER: US 07/681.492
CC FILING DATE: 04-APR-1991
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: WO PCT/EP90/01275
CC FILING DATE: 01-AUG-1990
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: EP 89 402 224.3
CC FILING DATE: 04-AUG-1989
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Schulman, Robert M.
CC REGISTRATION NUMBER: 31.196
CC REFERENCE/DOCKET NUMBER: 010830-027
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703) 836-6620
CC TELEFAX: (703) 836-2021
CC INFORMATION FOR SEQ ID NO: 4:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1046 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: double
CC TOPOLOGY: linear
CC MOLECULE TYPE: cDNA to mRNA
CC IMMEDIATE SOURCE:
CC CLONE: 3C9
CC
SQ SEQUENCE 1046 BP: 360 A; 254 C; 166 G; 266 T; 0 OTHER.

```

RESULT 12
ID PCT-US96-12860-1 STANDARD; DNA; UNC; 2589 BP.
AC xxxxxx
DT
DE Sequence 1, Application PC/TUS9612860
CC Sequence 1, Application PC/TUS9612860
CC GENERAL INFORMATION:
CC APPLICANT: TULARIK, INC.
CC TITLE OF INVENTION: INHIBITORS OF APOPTOSIS
CC NUMBER OF SEQUENCES: 14
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
CC STREET: 4 Embarcadero Center, Suite 3400
CC CITY: San Francisco
CC STATE: California
CC COUNTRY: USA
CC ZIP: 94111
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US96/12860.
CC FILING DATE: 06 AUG 1996
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: U.S. Serial Nos. 08/512,946 & 08/569,749
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Brezner, David J.
CC REGISTRATION NUMBER: 24,774
CC REFERENCE/DOCKET NUMBER: A-62464/DJB
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415)781-1989
CC TELEFAX: (415)398-3249
CC INFORMATION FOR SEQ ID NO: 1:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 2589 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: cDNA
CC SEQUENCE 2589 BP; 844 A; 436 C; 522 G; 787 T; 0 OTHER.
CC
CC Query Match 4.7%; Score 41; DB 2; Length 2589;
CC Best Local Similarity 83.6%; Pred. No. 2.13e-07;
CC Matches 51; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
DB 2510 AAACATCTCATTAAGTCGCTTTTAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 2569
Qy 805 aaataaataaagggtattatttcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 864
DB 2570 A 2570
Qy 865 a 865
RESULT 13
ID US-08-199-776-1 STANDARD; DNA; UNC; 3933 BP.
AC xxxxxx
DT
DE Sequence 1, Application US/08199776
CC Sequence 1, Application US/08199776
CC Patent No. 5594120
CC GENERAL INFORMATION:
CC APPLICANT: Brenner, Michael B.
CC APPLICANT: Parker, Christina M.
CC TITLE OF INVENTION: No. 5594120el integrin alpha subunit
CC NUMBER OF SEQUENCES: 25
CC CORRESPONDENCE ADDRESS:

```

COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/02044
FILING DATE: herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/199,776
FILING DATE: 18 February 1994
ATTORNEY/AGENT INFORMATION:
NAME: Plumer, Elizabeth R.
REGISTRATION NUMBER: 36,637
REFERENCE/DOCKET NUMBER: B0801/7020
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3933 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CELL TYPE: mucosal lymphocyte
FEATURE:
NAME/KEY: CDS
LOCATION: 126..3662
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 180..3659
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 126..1179
SEQUENCE 3933 BP: 938 A, 1062 C, 1074 G, 859 T, 0 OT

Query Match 4.7%; Score 41; DB 2; Length 3933;
Best Local Similarity 78.9%; Pred. No. 2.13e-07;
Matches 56; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

384	ATTGGTATTAAATAAATTACACITTTCTTTCGAAAAA	AAAAAAAAAAAAA	3900
Qy			
795	atttggtgcaataaatacaataaggttatttataaaaaa	aaaaaaaaaaaaa	854
Db	3907	AAAAAAAAAAAA	3917
Qy	855	aaaaaaaaaaaa	865

RESULT 15
ID US-07-642-002-1 STANDARD; DNA; UNC; 5108 BP.
AC xxxxxx

DE	Sequence 1, Application US/07642002
CC	Sequence 1, Application US/07642002
CC	Patent No. 5268465
CC	GENERAL INFORMATION:
CC	APPLICANT: Bredt, David S.
CC	APPLICANT: Hwang, Paul M.
CC	APPLICANT: Reed, Randall
CC	APPLICANT: Snyder, Solomon H.
CC	TITLE OF INVENTION: Purification and Molecular Cloning of Nitric
CC	TITLE OF INVENTION: Oxide Synthase
CC	NUMBER OF SEQUENCES: 2
CC	CORRESPONDENCE ADDRESS:

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	140	14.3	317	W03519	Prostate carcinoma tu	2.28e-03
2	109	11.1	316	13 W75702	Rat galectin-8.	7.19e-01
3	92	9.4	278	13 W72597	Osteragata proteince	1.41e-01
4	89	9.1	323	23 W46883	Protein sequence enco	2.34e-01
5	89	9.1	323	23 W11841	Human galectin-4-like	2.34e-01
6	89	9.1	323	20 W06551	Human collagen specific	2.34e-01
7	89	9.1	343	26 W36070	E. coli DNA polymeras	2.34e-01
8	89	9.1	983	13 R75761	BAV3 ORF6-encoded pr	2.34e-01
9	89	8.9	343	8 R40135	DNA polymerase III ho	3.27e-01
10	85	8.7	311	28 W36139	Snpr activator protei	4.56e-01
11	85	8.7	495	2 R07143	Neuronal nicotinic ac	5.38e-01
12	84	8.6	212	10 P54922	Cd11 polypeptide.	5.38e-01
13	84	8.6	704	19 W01102	Signal transducer and	5.38e-01
14	84	8.6	748	19 W01101	Signal transducer and	5.38e-01
15	83	8.5	264	3 R12531	Mac2.16 expression pr	6.34e-01
16	83	8.5	278	3 R13388	Mac-2 protein includi	6.34e-01
17	83	8.5	687	23 W22458	Yeast wild-type aceto	6.34e-01
18	83	8.5	687	1 P81152	Herbicide sensitive.	6.34e-01

```

QY 61 AWIS--R--WGQKLIAPFLYPQRFVLLLFQEGGLKALNG-QGLGATSMNQALE 115
Db 299 sdtleingdihlehrs 316
QY 116 QLERLISGSVQLYCVHS 133

RESULT 2
ID R75702 standard; Protein: 316 AA.
AC R75702;
DE 14-NOV-1995 (first entry)
KW Rat galectin-8.
OS Galectin-8; lectin; antiproliferative; antitumor.
FH Key
FT domain
FT 1..152
FT /label= CBD
FT /note= "N-terminal carbohydrate recognition
FT domain"
FT 153..174
FT /label= Linker_peptide
FT 175..316
FT /label= CBD
FT /note= "C-terminal carbohydrate recognition
FT domain"
FN W09515175-A.
PD 08-JUN-1995.
PF 05-DEC-1994; U13679.
PR 05-DEC-1993; IL-107880.
PA (RYCU/) RYCUS A.
PA (YEDA) YEDA RES & DEV CO LTD.
PI Zick Y;
DR WPI; 95-215153/28.
DR N-PSDB; Q90650.
DE Galectin-8 and corresp. DNA - used in regulation of cell growth, esp.
PT inhibition of cell proliferation
PS Claim 1; Fig. 1; 4pp; English.
CC cDNA encoding galectin-8, a novel 35 kDa S-type lectin, was cloned
CC when a lambda ZAP II rat liver cDNA library was screened with
CC antibodies directed against the C-terminal end of IRS-1.
CC Recombinant galectin-8 was expressed in E. coli and CHO cells.
SQ Sequence 316 AA;

Query Match 11.1%; Score 109; DB 13; Length 316;
Best Local Similarity 22.1%; Pred. No. 7.19e+01;
Matches 30; Conservative 41; Mismatches 58; Indels 7; Gaps 5;

Db 181 hlslfearlnasmgprtvvkvgevtntatsfnvdlvagrdsdialhlnprlnvkafr 240
QY 4 RLEVPCHALPQGLSPGQVIIVRGVLQEPKHTVSL-RDQAAHAPVTLRASFAADRTLAW 62
Db 241 nsflqdwgeeeenitcfofssgmymfemiiydvrefkvavngvhsleykhrfkd-lssi 299
QY 63 IS--R--WGQKLIAPFLYPQRFVLLLFQEGGLKALNG-QGLGATSMNQALEQL 117
Db 300 dclavdgdlrldvrs 315
QY 118 RELRISGSVQLYCVHS 133

RESULT 3
ID R72597 standard; Protein: 278 AA.
AC R72597;
DE 28-SEP-1995 (first entry)
KW Ostertagia protective antigen.
KW Protective antigen; parasite; immunogen; vaccine; lectin;
KW beta-galactoside binding protein.
OS Ostertagia circumcincta.
PD W09509182-A.
PN 06-APR-1995.
PF 27-SEP-1994; AU0573.
PR 28-SEP-1993; AU-001494.
PA (MEAT-) MEAT RES CORP.

PA (DYNE ) UNIV MELBOURNE.
PI Ashman K, Meeusen ENT, Newton SE, Walker J;
DR WPI; 95-147391/19.
DR N-PSDB; Q86743.
PT Protective antigens against Ostertagia, Trichostrongylus and
PT Fasciola species - are isolated by a method involving the use of
PT antibody probes from immune, challenged animals
PS Disclosure; Fig.19; 5pp; English.
CC A cDNA library prepared from 3rd stage larvae of O. circumcincta was
CC screened with a MAB raised against a 32-36 'doublet' antigen
CC identified by Western blotting. Clone specificity was determined by
CC plaque immunoassays, and clones were rescued into the plasmid form
CC (pMOSELOX) by plating on E. coli BM25.8 cells. Clones 3-2 and 5-2b
CC contained identical DNA sequences (given in Q86743) encoding a
CC lectin-like beta-galactoside-binding protein (R72597).
SQ Sequence 278 AA;

Query Match 9.4%; Score 92; DB 13; Length 278;
Best Local Similarity 28.2%; Pred. No. 1.41e+01;
Matches 11; Conservative 14; Mismatches 14; Indels 0; Gaps 0;

Db 9 ipyrskltepfepgqtlvtkvgtgedsvrftlnlhssa 47
QY 7 VPCSHALPQGLSPGQVIIVRGVLQEPKHTVSLRDQAA 45

RESULT 4
ID W46883 standard; Protein: 323 AA.
AC W46883;
DE 22-JUN-1998 (first entry)
DE Protein sequence encoded by a colon-specific gene.
KW Colon-specific gene; probe; detection; expression; human;
KW diagnostic assay; colon cancer; antibody; screening.
OS Homo sapiens.
PN US5733748-A.
PD 31-MAR-1998.
PF 06-JUN-1995; 469667.
PR 06-JUN-1995; US-469667.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen C, Yu G;
DR WPI; 98-229823/20.
DR N-PSDB; V16676.
PT Colon-specific nucleic acids - useful as probes for detecting colon
PT cancer micrometastases
PS Claim 1; Fig 9; 5pp; English.
CC W46876-80 and W4682-85 represent proteins encoded by colon-specific
CC genes. The polynucleotides encoding these proteins can be used
CC as probes to detect expression of the corresponding human genes,
CC e.g. in diagnostic assays for detecting micrometastases of colon cancer.
CC Recombinant cells containing the polynucleotides can be used to
CC produce the proteins, in order that antibodies can be raised and
CC used in further screening or diagnostics.
SQ Sequence 323 AA;

Query Match 9.1%; Score 89; DB 29; Length 323;
Best Local Similarity 24.8%; Pred. No. 2.34e+01;
Matches 33; Conservative 36; Mismatches 55; Indels 9; Gaps 5;

Db 192 vpyfgrlqggltarrtllikgyvptgktsfainfkvsggdialhlnprmgntvvrns 251
QY 7 VPCSHALPQGLSPGQVIIVRGVLQEPKHTVSLR-DQAAHAPVTLRASFAADRTLAW 62
Db 252 lngswgseekkithnpgf-pggffdlrcgldrfkvyangghlfdahrlsafqrvt 309
QY 63 IS-RWG--QKALISAPFLYPQRFVLLLFQEGGLKALNGQGLGATSMNQALEQLRE 119
Db 310 leiqgdvltlsyvg 322
QY 120 LRLSGSVQLYCVH 132

RESULT 5
ID W11841 standard; Protein: 323 AA.

```

W P E R E H (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Dec 15 15:56:27 1998; MasPar time 2.20 Seconds
Tabular output not generated. 427.387 Million cell updates/sec

Title: >US-09-109-864-2
Description: (1-133) from US09109864.pep
Perfect Score: 980
Sequence: 1 MSPRLVPCSHALPQGLSPG.....LEQLRELIRSGVOLYCVHS 133

Scoring table: PAM 150
Gap 11

Searched: 77309 seqs, 7078906 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: a-issued
1:5_COMB 2:PCT9_COMB 3:backfiles1

Statistics: Mean 28.678; Variance 129.097; scale 0.222

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description	Pred. No.
1	89	9.1	323	2	PCT-US95-0 Sequence 16, Applicati	1.07e+01
2	89	9.1	323	1	US-08-469- Sequence 16, Applicati	1.07e+01
3	89	9.1	343	1	US-08-279- Sequence 9, Applicati	1.07e+01
4	84	8.6	212	1	US-08-461- Sequence 35, Applicati	2.43e+01
5	84	8.6	748	1	US-08-839- Sequence 2, Applicati	2.43e+01
6	84	8.6	748	1	US-08-408- Sequence 2, Applicati	2.43e+01
7	81	8.3	847	1	US-08-781- Sequence 2, Applicati	3.94e+01
8	81	8.3	847	1	US-08-276- Sequence 2, Applicati	3.94e+01
9	80	8.2	785	1	US-08-526- Sequence 4, Applicati	4.62e+01
10	79	8.1	227	1	US-08-414- Sequence 4, Applicati	5.42e+01
11	79	8.1	357	2	PCT-US94-0 Sequence 12, Applicati	5.42e+01
12	79	8.1	357	1	US-08-145- Sequence 12, Applicati	5.42e+01
13	79	8.1	921	1	US-08-396- Sequence 2, Applicati	5.42e+01
14	79	8.1	921	1	US-08-818- Sequence 2, Applicati	5.42e+01
15	78	8.0	119	1	US-08-207- Sequence 20, Applicati	6.35e+01
16	78	8.0	125	1	US-08-313- Sequence 54, Applicati	6.35e+01
17	78	8.0	195	2	PCT-US93-0 Sequence 9, Applicati	6.35e+01
18	78	8.0	195	1	US-08-063- Sequence 9, Applicati	6.35e+01
19	78	8.0	677	1	US-08-646- Sequence 13, Applicati	6.35e+01
20	78	8.0	677	1	US-08-188- Sequence 13, Applicati	6.35e+01
21	77	7.9	312	1	US-08-414- Sequence 2, Applicati	7.44e+01
22	77	7.9	708	1	US-07-797- Sequence 2, Applicati	7.44e+01
23	77	7.9	708	1	US-08-308- Sequence 2, Applicati	7.44e+01

24 77 7.9 708 2 PCT-US95-0 Sequence 2, Applicatio 7.44e+01
25 77 7.9 973 1 US-08-162- Sequence 8, Applicatio 7.44e+01
26 76 7.8 70 1 US-08-476- Sequence 25, Applicati 8.70e+01
27 76 7.8 70 1 US-08-485- Sequence 25, Applicati 8.70e+01
28 76 7.8 70 1 US-07-956- Sequence 25, Applicati 8.70e+01
29 76 7.8 385 1 US-08-036- Sequence 5, Applicatio 8.70e+01
30 76 7.8 748 1 US-08-369- Sequence 10, Applicati 8.70e+01
31 76 7.8 748 2 PCT-US95-1 Sequence 10, Applicati 8.70e+01
32 76 7.8 749 1 US-08-781- Sequence 15, Applicati 8.70e+01
33 76 7.8 749 1 US-08-276- Sequence 15, Applicati 8.70e+01
34 76 7.8 905 1 US-08-072- Sequence 2, Applicatio 8.70e+01
35 76 7.8 3038 1 US-08-450- Sequence 2, Applicatio 8.70e+01
36 75 7.7 82 1 US-08-461- Sequence 11, Applicati 1.02e+02
37 75 7.7 82 2 PCT-US93-1 Sequence 11, Applicati 1.02e+02
38 75 7.7 106 1 US-08-643- Sequence 14, Applicati 1.02e+02
39 75 7.7 106 1 US-08-335- Sequence 45, Applicati 1.02e+02
40 75 7.7 106 1 US-08-406- Sequence 14, Applicati 1.02e+02
41 75 7.7 263 1 US-08-781- Sequence 17, Applicati 1.02e+02
42 75 7.7 372 1 US-08-278- Sequence 33, Applicati 1.02e+02
43 75 7.7 372 1 US-08-643- Sequence 33, Applicati 1.02e+02
44 75 7.7 983 1 US-08-162- Sequence 16, Applicati 1.02e+02
45 75 7.7 983 1 US-08-167- Sequence 10, Applicati 1.02e+02

ALIGNMENTS

RESULT 1
ID PCT-US95-07289-16 STANDARD; PRT; 323 AA.
XX
AC xxxxxx
XX
DT
XX
XX
DE Sequence 16, Application PC/TUS9507289
XX
XX
CC Sequence 16, Application PC/TUS9507289
CC GENERAL INFORMATION:
CC APPLICANT: Yu, Guo-Liang
CC APPLICANT: Rosen, Craig
CC TITLE OF INVENTION: Colon Specific Genes and Proteins
CC NUMBER OF SEQUENCES: 24
CC CORRESPONDENCE ADDRESS:
CC ADDRESS: Carella, Byrne, Bain, Gilfillan, Cecchi,
CC ADDRESSEE: Stewart & Oistein
CC STREET: 6 Becker Farm Road
CC CITY: Roseland
CC STATE: NJ
CC COUNTRY: USA
CC ZIP: 07068-1739
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: IBM PC compatible
CC SOFTWARE: PC-DOS/MS-DOS
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/07289
CC FILING DATE: 06-JUN-1995
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Ferraro, Gregory D.
CC REGISTRATION NUMBER: 36,134
CC REFERENCE/DOCKET NUMBER: 325800-265
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 201-994-1700
CC TELEFAX: 201-994-1744
CC INFORMATION FOR SEQ ID NO: 16:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 323 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 323 AA: 35941 MW; 553806 CN;

CC	NUMBER OF SEQUENCES: 4	
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESSEE: Flehr, Hohnbach, Test, Albritton & Herber	
CC	STREET: 850 Hansen Way, #200	
CC	CITY: Palo Alto	
CC	STATE: CA	
CC	COUNTRY: USA	
CC	.ZIP: 94304	
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE: Floppy disk	
CC	COMPUTER: IBM PC compatible	
CC	OPERATING SYSTEM: PC-DOS/MS-DOS	
CC	SOFTWARE: Patentin Release #1.0, Version #1.30	
CC	CURRENT APPLICATION DATA:	
CC	APPLICATION NUMBER: US/08/839,164	
CC	FILING DATE: 23-APR-1997	
CC	CLASSIFICATION: 536	
CC	PRIOR APPLICATION DATA:	
CC	APPLICATION NUMBER: 08/408,318	
CC	FILING DATE:	
CC	ATTORNEY/AGENT INFORMATION:	
CC	NAME: Osman, Richard A.	
CC	REGISTRATION NUMBER: 36,627	
CC	REFERENCE/DOCKET NUMBER: A-60845	
CC	TELECOMMUNICATION INFORMATION:	
CC	TELEPHONE: 415-494-8700	
CC	TELEFAX: 415-494-8771	
CC	INFORMATION FOR SEQ ID NO: 2:	
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH: 748 amino acids	
CC	TYPE: amino acid	
CC	TOPOLOGY: linear	
CC	MOLECULE TYPE: protein	
CC	SEQUENCE 748 AA; 85940 MW; 3117180 CN;	
CC	Query Match 8.6%; Score 84; DB 1; Length 748;	
CC	Best Local Similarity 23.8%; Pred. No. 2.43e+01;	
CC	Matches 19; Conservative 27; Mismatches 30; Indels 4; Gaps 4;	
DB	522 LTQSYSDGHLTW-AKFCHEHLPGKSFYFWT-WLAAILDLKKHLPLWIDYVNGFVS 579	
QY	49 VTLRASFADTLRWISWGOKKLISAPFLFYQRFVEVLFFOEGGKILALNGGILGATS 108	
DB	580 KEKERL-LLKD-KMPGTFL 597	
QY	109 MNQOAEQLRELRLSGSVQL 128	
RESULT	6	
US	US-08-408-318-2	STANDARD; PRT; 748 AA.
XX	XXXXXX	
AC		
XX		
XX		
DT		
XX		
DE		
XX		
CC	Sequence 2, Application US/08408318	
CC	Sequence 2, Application US/08408318	
CC	Patent No. 5639858	
CC	GENERAL INFORMATION:	
CC	APPLICANT: Hoey, Timothy	
CC	TITLE OF INVENTION: Human Signal Transducers and Binding	
CC	TITLE OF INVENTION: Assays	
CC	NUMBER OF SEQUENCES: 4	
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESSEE: Flehr, Hohnbach, Test, Albritton & Herber	
CC	STREET: 850 Hansen Way, #200	
CC	CITY: Palo Alto	
CC	STATE: CA	
CC	COUNTRY: USA	
CC	.ZIP: 94304	
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE: Floppy disk	

CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/408.318
CC FILING DATE:
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Osman, Richard A
CC REGISTRATION NUMBER: 36,627
CC REFERENCE/DOCKET NUMBER: A-60845
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 415-494-8700
CC TELEFAX: 415-494-8771
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 748 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE: 748 AA; 85940 MW; 3117180 CN;
SQ
Query Match 8.6%; Score 84; DB 1; Length 748;
Best Local Similarity 23.8%; Pred. No. 2.43e+01;
Matches 19; Conservative 27; Mismatches 30; Indels 4; Gaps 4;
Db 522 LTVOSYSDGHLTW-AKFCHELPKSTFTWT-WLEALDLIKHLPWIDGYMGFVS 579
QY 49 VTLRASPADRTLWISRGQKLLISAPFLFYQRFVEVLLLFQEGGLKALNGQIGATS 108
Db 580 KEREL-LLKD-KMPGTFL 597
QY 109 MNQALRELIRISGVOL 128
RESULT 7
ID US-08-781-890-2 STANDARD; PRT; 847 AA.
AC xxxxxx
XX
XX
XX
XX
DE Sequence 2, Application US/08/781890
XX
Sequence 2, Application US/08/781890
Patent No. 5710266
CC GENERAL INFORMATION:
CC APPLICANT: McKnight, Steven L
CC TITLE OF INVENTION: INTERLEUKIN-4 SIGNAL TRANSDUCERS AND
CC TITLE OF INVENTION: BINDING ASSAYS
CC NUMBER OF SEQUENCES: 17
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
CC STREET: 4 Embarcadero Center, Suite 3400
CC CITY: San Francisco
CC STATE: California
CC COUNTRY: USA
CC ZIP: 94111-4187
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/781.890
CC FILING DATE: 05-JAN-1997
CC CLASSIFICATION: 536
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/276.099
CC FILING DATE: 15-JUL-1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Osman, Richard Aron

CC REGISTRATION NUMBER: 36,627
CC REFERENCE/DOCKET NUMBER: A-59451-1/RAO
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 781-1989
CC TELEFAX: (415) 398-3249
CC TELEX: 910 277299
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 847 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE: 847 AA; 94134 MW; 3802900 CN;
SQ
Query Match 8.3%; Score 81; DB 1; Length 847;
Best Local Similarity 29.6%; Pred. No. 3.94e+01;
Matches 21; Conservative 16; Mismatches 26; Indels 8; Gaps 7;
Db 461 AEVGTNRGL-L--PEBF-LFLAQKIFNDASLSM-EAFORHSVSW-SQFNKEILLGRGTF 514
QY 20 GQVIVRGVLQEPKFTVSLRQQA-AHAPVTLRASFADETLAWISRGQKLLISAPFL 78
Db 515 W-QWEDGVLDL 524
QY 79 YPQRFVEVLL 89
RESULT 8
ID US-08-276-099A-2 STANDARD; PRT; 847 AA.
AC xxxxxx
XX
XX
XX
XX
DE Sequence 2, Application US/08/276099A
XX
Sequence 2, Application US/08/276099A
Patent No. 5591825
CC GENERAL INFORMATION:
CC APPLICANT: McKnight, Steven L
CC TITLE OF INVENTION: INTERLEUKIN-4 SIGNAL TRANSDUCERS AND
CC TITLE OF INVENTION: BINDING ASSAYS
CC NUMBER OF SEQUENCES: 17
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
CC STREET: 4 Embarcadero Center, Suite 3400
CC CITY: San Francisco
CC STATE: California
CC COUNTRY: USA
CC ZIP: 94111-4187
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/276.099A
CC FILING DATE: 15-JUL-1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Osman, Richard Aron
CC REGISTRATION NUMBER: 36,627
CC REFERENCE/DOCKET NUMBER: A-59451-1/RAO
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 781-1989
CC TELEFAX: (415) 398-3249
CC TELEX: 910 277299
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 847 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear

Db 80 MA-RLTNSCKECLGFSLTPSHLGIVINAYVLEEVN 114
|: || :| | :|: ||: ||: |

CC	APPLICANT:	Hogan, Patrick Gerald
CC	APPLICANT:	McCaifrey, Patricia
CC	APPLICANT:	Jain, Jugnu
CC	TITLE OF INVENTION:	NF-ATp, A T LYMPHOCTYTE
CC	TITLE OF INVENTION:	DNA-BINDING PROTEIN
CC	NUMBER OF SEQUENCES:	22
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESSEE:	Fish & Richardson
CC	STREET:	225 Franklin Street
CC	CITY:	Boston
CC	STATE:	Massachusetts
CC	COUNTRY:	U.S.A.
CC	ZIP:	02110-2804
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE:	3.5" Diskette, 1.44 Mb
CC	COMPUTER:	IBM PS/2 Model 50Z or 55SX
CC	OPERATING SYSTEM:	MS-DOS (Version 5.0)
CC	SOFTWARE:	WordPerfect (Version 5.1)
CC	CURRENT APPLICATION DATA:	
CC	APPLICATION NUMBER:	PC/US94/00545
CC	FILING DATE:	18-JAN-1994
CC	CLASSIFICATION:	
CC	PRIOR APPLICATION DATA:	
CC	APPLICATION NUMBER:	08/145,006
CC	FILING DATE:	October 29, 1993
CC	APPLICATION NUMBER:	08/017,052
CC	FILING DATE:	February 11, 1993
CC	APPLICATION NUMBER:	08/006,067
CC	FILING DATE:	January 15, 1993
CC	ATTORNEY/AGENT INFORMATION:	
CC	NAME:	Fraser, Janis K.
CC	REGISTRATION NUMBER:	34,819
CC	REFERENCE/DOCKET NUMBER:	04590/007001
CC	TELECOMMUNICATION INFORMATION:	
CC	TELEPHONE:	(617) 542-5070
CC	TELEFAX:	(617) 542-8906
CC	TELEX:	200154
CC	INFORMATION FOR SEQ ID NO:	12:
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH:	357
CC	TYPE:	amino acid
CC	STRANDEDNESS:	
CC	TOPOLOGY:	linear
CC	SEQUENCE	357 AA; 37623 MW; 700002 CN;
Db	Query Match	8.1%; Score 79; DB 2; L
Qy	Best Local Similarity	36.4%; Pred. No. 5.42e+01;
	Matches	12; Conservative 12; Mismatches 8
Db	51	LSPRIETPSHILQAVGLMRDA-GLLVSEP 82
Qy	1	MSRFLVPCSHALPQGLSPGVIVRGLVLEP 33
		: : : :
RESLT	12	
ID	DS-08-145-006C-12	STANDARD; PRU; 357
XX	XXXXXX	
AC		
XX		
DT		
XX		
DE		
XX		
CC	Sequence 12, Application US/08145006C	
CC	Sequence 12, Application US/08145006C	
CC	Patent No. 5656452	
CC	GENERAL INFORMATION:	
CC	APPLICANT:	Rao, Anjana
CC	APPLICANT:	Hogan, Patrick Gerald
CC	APPLICANT:	McCaifrey, Patricia
CC	APPLICANT:	Jain, Jugnu
CC	TITLE OF INVENTION:	NF-ATp, A T LYMPHOCTYTE
CC	TITLE OF INVENTION:	DNA-BINDING PROTEIN
CC	NUMBER OF SEQUENCES:	21

```

CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Fish & Richardson P.C.
CC STREET: 225 Franklin Street
CC CITY: Boston
CC STATE: Massachusetts
CC COUNTRY: U.S.A.
CC ZIP: 02110-2804
CC
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
CC COMPUTER: IBM PS/2 Model 502 or 55SX
CC OPERATING SYSTEM: MS-DOS (Version 5.0)
CC SOFTWARE: WordPerfect (Version 5.1)
CC
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/145.006C
CC FILING DATE: October 29, 1993
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/017,052
CC FILING DATE: February 11, 1993
CC APPLICATION NUMBER: 08/006,067
CC FILING DATE: January 15, 1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Fraser, Janis K.
CC REGISTRATION NUMBER: 34,819
CC REFERENCE/DOCKET NUMBER: 04590/007001
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 542-5070
CC TELEFAX: (617) 542-8906
CC TELEX: 200154
CC
CC INFORMATION FOR SEQ ID NO: 12:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 357
CC TYPE: amino acid
CC STRANDEDNESS:
CC TOPOLOGY: linear
CC
CC SEQ SEQUENCE 357 AA; 37623 MW; 700002 CN;
CC
CC Query Match 8.1%; Score 79; DB 1: Length 357;
CC Best Local Similarity 36.4%; Pred. No. 5.42e+01;
CC Matches 12; Conservative 12; Mismatches 8; Indels
CC
CC Db 51 LSPRIETTPSHLIQAVGPLRMRA-GLIVEQP 82
CC :|||: || |::|:: ||::|
CC 1 MSPREVPCHALPQGLSPGVIIVRGLVLP 33
CC
CC RESULT 13
CC ID US-08-396-479B-2 STANDARD; PRT; 921 AA.
CC XX XXXXXX
CC
CC DT
CC XX
CC DE
CC XX Sequence 2, Application US/08396479B
CC XX Sequence 2, Application US/08396479B
CC Patent No. 5612455
CC
CC GENERAL INFORMATION:
CC APPLICANT: HOEY, Timothy
CC TITLE OF INVENTION: NUCLEAR FACTORS AND BINDING ASSAY
CC NUMBER OF SEQUENCES: 18
CC
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: FLEHR, ROBBACH, TEST, ALBRITTON & HERBERT
CC STREET: 4 Embarcadero Center, Suite 3400
CC CITY: San Francisco
CC STATE: California
CC COUNTRY: USA
CC ZIP: 94111
CC
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0. Version #1.30

```


UU

Db 67 RVTMLRDTSKNOFSRLSSVTAAADTAIYICARHGDDP-AWFAYWGOGSLVTVS 118
QY 26 RGLVLOE-PK-HFTVSLRD-QAAHAPVTILRASFAADRTLAWISRWGOKKLISAP 75

Search completed: Tue Dec 15 15:56:35 1998
Job time : 8 secs.

Result No.	Query		Score	Match	Length	DB	ID	Description	Pred. No.
1	174	17.8	145	2	A55932			galactin-5 - rat	7.78e-12
2	140	13.3	317	2	JC6147			prostate carcinoma tu	1.19e-06
3	136	13.9	136	2	155459			galactin-7 - human	4.56e-06
4	112	11.4	279	2	A42846			beta-galactoside-bind	9.92e-03
5	109	11.1	316	2	A55975			galactin-8 - rat	2.46e-02
6	109	11.1	324	2	A46631			lactose-binding lecti	2.46e-02
7	106	10.8	448	2	A27522			UDP-N-acetylglucosami	6.03e-02
8	99	10.1	135	1	LNC514			14K beta-galactoside-	4.60e-01
9	98	10.0	308	2	S13584			nosf protein - pseudo	6.11e-01
10	95	9.7	746	2	A40636			enterobactin receptor	1.41e+00
11	95	9.7	1321	2	A60185			sodium channel protei	1.41e+00
12	93	9.5	493	2	A65126			probable general secr	2.45e+00
13	93	9.5	1522	2	JC1101			sodium channel protei	2.45e+00
14	92	9.4	583	2	A29154			complement factor 1 (3.22e+00
15	92	9.4	620	1	OQEC80			glutathione-regulated	3.22e+00
16	91	9.3	703	2	B41538			ATP-binding cassette	4.21e+00
17	90	9.2	242	2	A49800			galactoside-binding p	5.50e+00
18	90	9.2	510	2	A64101			multidrug resistance	5.50e+00
19	89	9.2	761	2	S20458			pqgf protein - Klebsi	5.50e+00
20	89	9.1	225	2	A48640			global nitrogen regul	7.18e+00
21	89	9.1	833	2	JOL866			hypothetical 87.1k pr	7.18e+00
22	89	9.1	862	2	S57964			lipoxygenase (EC 1.13	7.18e+00
23	88	9.0	137	2	B40361			virC-region hypothei	9.34e+00

```

Db 70 NTQINNSGPEERSLPGSMPPSGQR-F-SYVILCEGHCPCVAVDGHQICVSHRLMLNLPD 128
QY 63 ---ISR-WG-QRKLISAPFLFY-PQRFFEVLLLFQEGGLKALNGGIGATSMNQQALEQ 116
Db 129 INTLEVAGDIQL 140
QY 117 LRELRISSGVQL 128

RESULT 2
ENTRY JG6147 #type complete
TITLE prostate carcinoma tumor antigen 1 - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 02-Sep-1997 #sequence_revision 03-Sep-1997 #text_change
ACCESSIONS JG6147
REFERENCE JG6147
#authors Su, Z.Z.; Lin, J.; Shen, R.; Fisher, P.E.; Goldstein, N.I.;
#journal Proc. Natl. Acad. Sci. U.S.A. (1996) 93:7252-7257
#title Surface-epitope masking and expression cloning identifies the
human prostate carcinoma tumor antigen gene PCTA-1 a member
of the galectin gene family.
#accession JG6147
#molecule_type mRNA
#residues 1-317 #label SUA
#cross-references GB:L78132; NID:g1932711; PID:g1932712
COMMENT This protein is a therapeutic reagent for intervention in pervasive
and fatal neoplastic disease. It is involved in an earlier
genetic change in human prostate cancer development. This protein
is a member of the galectin family.
KEYWORDS tumor
SUMMARY #length 317 #molecular-weight 35719 #checksum 1508

Query Match 14.38; Score 140; DB 2; Length 317;
Best Local Similarity 24.6%; Pred. No. 1.19e-06;
Matches 34; Conservative 41; Mismatches 56; Indels 7; Gaps 5;

Db 180 TPQLSLPFAARLTMPGRTVVQGVNANAKSFNVDLLAGKSKDIALHNPRLNIKAF 239
QY 2 SPRLVPCSHALPQGLSPGVIIIVRGLVQEPKHFVSL-RDQAHAAPVILRASFADERTL 60
Db 240 VRNSFTQESNGEERNITSPFPSPGMFYFEMIIYCDVREFKAVNGVHSLEYKHFEK-LS 298
QY 61 AWIS--R--WGQKLISAPFLFY-PQRFFEVLLLFQEGGLKALNG-QGLGATSMNQQALE 115
Db 299 SIDLTLINGDIHLEVR 316
QY 116 QLRRLRISGSVQLCVHS 133

RESULT 3
ENTRY I55469 #type complete
TITLE galectin-7 - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 24-May-1996 #sequence_revision 24-May-1996 #text_change
ACCESSIONS I55469
REFERENCE I55469
#authors Madsen, P.; Rasmussen, H.H.; Flint, T.; Gromov, P.; Kruse,
T.A.; Honore, B.; Vorum, H.; Celis, J.E.
#journal J. Biol. Chem. (1995) 270:5823-5829
#title Cloning, expression, and chromosome mapping of human
galectin-7.
#cross-references MUID:95197604
#accession I55469
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-136 #label RES
#cross-references GB:L07759; NID:g182131; PID:g182132
SUMMARY #length 136 #molecular-weight 15075 #checksum 8853

```

```

Query Match 13.9%; Score 136; DB 2; Length 136;
Best Local Similarity 28.1%; Pred. No. 4.55e-06;
Matches 38; Conservative 39; Mismatches 47; Indels 11; Gaps 6;

Db 3 NVPHKSSLPGRGTIVLRINGLYPPNARFHVNLCCGEGSDAALHFNPLDTSYEVF 62
QY 6 EVPCSHALPQGLSPGVIIIVRGLVQEPKHFVSL--RDQAHAAPVILRASFADERTLAW 62
Db 63 NSKEQGSWGREERGPGVFFQR-GQPF-EVLLIASDDGFK-AVVGDAQYBHFHRLPLARY 119
QY 63 ISR----WG-QRKLISAPFLFY-PQRFFEVLLLFQEGGLKALNGGIGATSMNQQALEQ 117
Db 120 RLVEVGDDVOLDVSR 134
QY 118 RELRISGSVQLCVH 132

RESULT 4
ENTRY A42846 #type complete
TITLE beta-galactoside-binding protein GBP - Caenorhabditis elegans
ALTERNATE_NAMES beta-galactoside-binding lectin homolog, 32K
ORGANISM #formal_name Caenorhabditis elegans
DATE 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change
ACCESSIONS A42846; S27798
REFERENCE A42846
#authors Hirabayashi, J.; Satoh, M.; Kasai, K.
#journal J. Biol. Chem. (1992) 267:15485-15490
#title Evidence that Caenorhabditis elegans 32-kDa
beta-galactoside-binding protein is homologous to
vertebrate beta-galactoside-binding lectins. cDNA cloning
and deduced amino acid sequence.
#cross-references MUID:92348399
#accession A42846
#status preliminary
#molecule_type mRNA
#residues 1-279 #label HIR
#cross-references EMBL:N94671; NID:g156209; PID:g156210
#note sequence extracted from NCBI backbone (NCBIN:109554)
SUMMARY #length 279 #molecular-weight 31809 #checksum 854

Query Match 11.4%; Score 112; DB 2; Length 279;
Best Local Similarity 35.9%; Pred. No. 9.92e-03;
Matches 14; Conservative 12; Mismatches 13; Indels 0; Gaps 0;

Db 11 VYRSVLQKFPQGTIVKGTIDESQRTINLHSTTA 49
QY 7 VPCSHALPQGLSPGVIIIVRGLVQEPKHFVSLRDQA 45

RESULT 5
ENTRY A55975 #type complete
TITLE galectin-8 - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 23-Mar-1995 #sequence_revision 05-Apr-1995 #text_change
ACCESSIONS A55975
REFERENCE A55975
#authors Hadari, Y.R.; Paz, K.; Dekel, R.; Mestrovic, T.; Accili, D.;
Zick, Y.
#journal J. Biol. Chem. (1995) 270:3447-3453
#title Galectin-8. A new rat lectin, related to galectin-4.
#accession A55975
#status preliminary
#molecule_type mRNA
#residues 1-316 #label HAD
#cross-references GB:U09824; NID:g171031; PID:g171032
SUMMARY #length 316 #molecular-weight 36038 #checksum 9989

Query Match 11.1%; Score 109; DB 2; Length 316;
Best Local Similarity 22.1%; Pred. No. 2.46e-02;
Matches 30; Conservative 41; Mismatches 58; Indels 7; Gaps 5;

```

[illegible]

	beta-galactoside-binding lectin of chick embryo.
#cross-references	MUID:87250364
#accession	JT0009
##molecule_type	protein
##residues	2-110,112-135 ##label HIR
REFERENCE	AZ4062
#authors	Ohyama, Y.; Hirabayashi, J.; Oda, Y.; Ohno, S.; Kawasaki, H.; Suzuki, K.; Kasai, K.
#journal	Biochem. Biophys. Res. Commun. (1986) 134:51-56
#title	Nucleotide sequence of chick 14k beta-galactoside-binding lectin mRNA.
#cross-references	MUID:86130505
#accession	AZ4062
##molecule_type	mRNA
##residues	12-135 ##label OH2
GENETICS	
#introns	3/3; 30/2; 118/3
CLASSIFICATION	#superfamily beta-galactoside-binding lectin
KEYWORDS	acetylated amino end; lectin
FEATURE	
2-135	
#product	beta-galactoside-binding lectin #status
#experimental	#label MAY
#modified_site	acetylated amino end (Ser) (in mature form) #status experimental
SUMMARY	#length 135 #molecular-weight 15063 #checksum 8342
Query Match	10.1%; Score 99; DB 1; Length 135;
Best Local Similarity	27.9%; Pred. No. 4.60e-01;
Matches	12; Conservative 15; Mismatches 16; Indels 0; Gaps 0;
Db	12 GLRPGORLVKGIAPNAKSFVMMGLGKDTHLGHFNPRFDAH 54 : : : : : : : :
Oy	16 GLSPGOVIIVRGVLQEPKHFTVSLRDQAHAHVTLRASFAADR 58 : : : : : : : :
RESULT	9
ENTRY	S13584
TITLE	nosf protein - Pseudomonas stutzeri
ORGANISM	#formal_name Pseudomonas stutzeri
DATE	21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
20-Feb-1998	
ACCESSIONS	S13584
REFERENCE	S13582
#authors	Zumft, W.G.; Viebrock-Sambale, A.; Braun, C.
#journal	Eur. J. Biochem. (1990) 192:591-599
#title	Nitrous oxide reductase from denitrifying Pseudomonas stutzeri. Genes for copper-processing and properties of the deduced products, including a new member of the family of ATP/GTP-binding proteins.
#cross-references	MUID:91006150
#accession	S13584
##status	preliminary
##molecule_type	DNA
##residues	1-308 ##label ZDM
##cross-references	EMBL:X33676; NID:g45844; PID:g45849
CLASSIFICATION	#superfamily ATP-binding cassette homology
FEATURE	
19-210	
#domain	ATP-binding cassette homology #label ABC1
SUMMARY	#length 308 #molecular-weight 33777 #checksum 7279
Query Match	10.0%; Score 98; DB 2; Length 308;
Best Local Similarity	33.3%; Pred. No. 6.11e-01;
Matches	21; Conservative 15; Mismatches 20; Indels 7; Gaps 7;
Db	182 ILCSHVLPGVEAHNRAAILAKG-CLOAVGSLG-OLRAEAG-LPVRTFRASGISERD-SWL 237 : : : : : : : : :
Oy	7 VPCSHALFOGLSP-GOV-IIIVRGVLQEPKHFTVSLRDQAHAHVTLRASFAADR 63 : : : : : : : : :
Db	238 ORW 240
Oy	64 SRW 66

RESULT	10
ENTRY	A40636
TITLE	enterobactin receptor protein, 80K - Pseudomonas aeruginosa
ORGANISM	#formal_name Pseudomonas aeruginosa
DATE	21-Sep-1993 #sequence_revision 18-Mar-1994 #text_change
20-Mar-1998	
ACCESSIONS	A40636
REFERENCE	A40636
#authors	Dean, C.R.; Poole, K.
#journal	J. Bacteriol. (1993) 175:317-324
#title	Cloning and characterization of the ferric enterobactin receptor gene (pfea) of Pseudomonas aeruginosa.
#cross-references	MUID:93123148
#contents	K407
#accession	A40636
##status	preliminary
##molecule_type	nucleic acid
##residues	1-746 ##label DEA
##cross-references	GB:M98033; MID:g151434; PID:g151435
##note	sequence extracted from NCBI backbone (NCBIN:L22308, NCBIP:L22309)
CLASSIFICATION	#superfamily ferriterocheilin receptor
SUMMARY	#length 746 #molecular-weight 80967 #checksum 8531
Query Match	9.7%; Score 95; DB 2; Length 746;
Best Local Similarity	37.8%; Pred. NO. 1.41e+00;
Matches	14; Conservative 10; Mismatches 12; Indels 1; Gaps 1;
Db	363 YTATLRDLTAGENVLPRLHGVEQTFLILGSEWTOKL 399 : : : : : : : :
Oy	36 FTVSLRDQAHAHVTLRASFA-DRTLAWISRWGOKKL 71 : : : : : : : :
RESULT	11
ENTRY	A60165
TITLE	sodium channel protein - fruit fly (Drosophila melanogaster (fragments))
ORGANISM	#formal_name Drosophila melanogaster
DATE	22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change
16-Feb-1997	
ACCESSIONS	S04029; A60165
REFERENCE	S04029
#authors	Salkoff, L.; Butler, A.; Scavarda, N.; Wei, A.
#journal	Nucleic Acids Res. (1987) 15:8569-8572
#title	Nucleotide sequence of the putative sodium channel gene from Drosophila: the four homologous domains.
#cross-references	MUID:88040482
#accession	S04029
##molecule_type	DNA
##residues	1-362; 363-626; 627-1321 ##label SAL
##cross-references	EMBL:X14394
REFERENCE	A60165
#authors	Salkoff, L.; Butler, A.; Wei, A.; Scavarda, N.; Giffen, K.; Ifune, C.; Goodman, R.; Mandel, G.
#journal	Science (1987) 237:744-749
#title	Genomic organization and deduced amino acid sequence of a putative sodium channel gene in Drosophila.
#cross-references	MUID:87292090
#accession	A60165
##status	nucleic acid sequence not shown
##molecule_type	DNA
##residues	40-355; 363-560.'P'.562-626; 632-1263 ##label SA2
##cross-references	EMBL:X14394
##note	part of this sequence was confirmed by mRNA sequencing the authors' translation is shown at position 561
GENETICS	
#gene	FlyBase:NACP60E
#cross-references	FlyBase:FBgn0002920
#introns	237/2; 310/3; 362/3; 414/3; 471/3; 531/3; 581/1; 626/3; 751/2; 801/1; 908/1; 960/3; 1005/1
SUMMARY	#length 1321 #checksum 3413
Query Match	9.7%; Score 95; DB 2; Length 1321;

```

113-134      #domain transmembrane #status predicted #label TM3\
144-167      #domain transmembrane #status predicted #label TM4\
180-201      #domain transmembrane #status predicted #label TM5\
309-331      #domain transmembrane #status predicted #label TM6\
407-426      #domain transmembrane #status predicted #label TM7\
443-464      #domain transmembrane #status predicted #label TM8\
473-491      #domain transmembrane #status predicted #label TM9\
499-522      #domain transmembrane #status predicted #label TM10\
532-553      #domain transmembrane #status predicted #label TM11\
605-625      #domain transmembrane #status predicted #label TM12\
778-797      #domain transmembrane #status predicted #label TM13\
816-837      #domain transmembrane #status predicted #label TM14\
847-868      #domain transmembrane #status predicted #label TM15\
916-937      #domain transmembrane #status predicted #label TM16\
1010-1032    #domain transmembrane #status predicted #label TM17\
1101-1120    #domain transmembrane #status predicted #label TM18\
1133-1154    #domain transmembrane #status predicted #label TM19\
1163-1184    #domain transmembrane #status predicted #label TM21\
1195-1218    #domain transmembrane #status predicted #label TM22\
1237-1258    #domain transmembrane #status predicted #label TM23\
1324-1346    #domain transmembrane #status predicted #label TM24\

SUMMARY      #length 1522 #molecular-weight 174113 #checksum 2729

Query Match      9.5% Score 93; DB 2; Length 1522;
Best Local Similarity 35.0%; pred. No. 2.45e+00;
Matches 14; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

Db 878 SLRTLRLPLRAVSRQGGKIIVNALMLSPISFNVLV 917
      :||: || | ||| :|| :||: |||:
Qy 50 TLRAFADRLTAWISRQGGKLIISAPFLYFORFEVILL 89

RESULT 14
ENTRY
TITLE      A29154      #type complete
ALTERNATE_NAMES  Complement factor I (EC 3.4.21.45) precursor - human
ORGANISM   C3B/C4b inactivator
#formal_name Homo sapiens #common_name man
DATE       31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change
          10-Oct-1997;
ACCESSIONS A29154; A28434; S66420
REFERENCE   A29154
#authors   Catterall, C.F.; Lyons, A.; Sim, R.B.; Day, A.J.; Harris,
          T.J.R.
#journal   Biochem. J. (1987) 242:849-856
#title     Characterization of the primary amino acid sequence of human
          complement control protein factor I from an analysis of
          cDNA clones.
#cross-references MIM:87241401
#accession A29154
#molecule_type mRNA
#residues  1-583 #label CAT
#cross-references GB:Y00318
REFERENCE   A28434
#authors   Goldberger, G.; Bruns, G.A.P.; Rits, M.; Edge, M.D.;
          Kwiatkowski, D.J.
#journal   J. Biol. Chem. (1987) 262:10065-10071
#title     Human complement factor I: analysis of cDNA-derived primary
          structure and assignment of its gene to chromosome 4.
#cross-references MIM:87280021
#accession A28434
#molecule_type mRNA
#residues  1-557, 'r', 559-583 #label GOL
#cross-references GB:J02770; NID:gl82606; PID:gl82607
REFERENCE   S66420
#authors   Ullman, C.G.; Haris, P.I.; Smith, K.F.; Sim, R.B.; Emery,
          V.C.; Perkins, S.J.
#journal   FEBS Lett. (1995) 371:199-203
#title     Beta-sheet secondary structure of an LDL receptor domain from
          complement factor I by consensus structure predictions and
          spectroscopy.
#accession S66420
#molecule_type protein

```

```

#residues      258-269  ##label ULL
GENETICS
#gene          GDB:IF
#cross-references GDB:120077; OMIM:217030
#map_position  4q24-q25
CLASSIFICATION #superfamily LDL receptor ligand-binding repeat homology;
               trypsin homology
KEYWORDS       glycoprotein; hydrolase; serine proteinase
FEATURE
1-18           #domain signal sequence #status predicted #label SIG\
19-335         #product complement factor I heavy chain #status
               #predicted #label CFH\
221-255        #domain LDL receptor ligand-binding repeat homology
               #label LDLA\
259-293        #domain LDL receptor ligand-binding repeat homology
               #label LDL2\
340-583        #product complement factor I light chain #status
               #predicted #label CFL\
340-569        #domain trypsin homology #label TRY\
70,103,177,464,494,
536            #binding_site carbohydrate (Asn) (covalent) #status
               predicted
SUMMARY        #length 583 #molecular-weight 65720 #checksum 3015
               9.4%; Score 92; DB 2; Length 583;
Query Match    Best Local Similarity 52.4%; Pred. No. 3.22e+00;
Matches 11; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
Db 485 QMGVKLISCKRYGNRYE 505
QY 65 RWGKKLISAPFLFPQRF 85
               ::::: ||||| |||||

RESULT 15
ENTRY    QOECRD      #type complete
TITLE    glutathione-regulated potassium efflux system protein kefc -
         Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE     05-Apr-1983 #sequence_revision 01-Mar-1996 #text_change
         14-Nov-1997
ACCESSION S40568; S15165; A04444; G64725
REFERENCE S40531
#authors   Yura, T.; Mori, H.; Nagai, H.; Nagata, T.; Ishihama, A.;
           Fujita, N.; Isono, K.; Mizobuchi, K.; Nakata, A.
#submission submitted to the EMBL Data Library, December 1992
#description Systematic sequencing of the Escherichia coli genome:
           analysis of the 0-2.4min region.
#accession S40568
#status    preliminary
#molecule_type DNA
#residues 1-620 ##label YUR
#cross-references EXBL:D10483; NID:g216434; PID:d1001795; PID:g216472
S15165
#authors   Munro, A.W.; Ritchie, G.Y.; Lamb, A.J.; Douglas, R.M.; Booth,
           I.R.
#journal    Mol. Microbiol. (1991) 5:607-616
#title      The cloning and DNA sequence of the gene for the
           glutathione-regulated potassium-efflux system kefc of
           Escherichia coli.
#cross-references MUID:9126044
#accession S15165
#molecule_type DNA
#residues 1-620 ##label MUN
#cross-references EXBL:X56742; NID:g41874; PID:g41875
A93704
#authors   Smith, D.R.; Calvo, J.M.
#journal    Nucleic Acids Res. (1980) 8:2255-2274
#title      Nucleotide sequence of the Escherichia coli gene coding for
           dihydrofolate reductase.
#cross-references MUID:81053692
#accession A04444
#molecule_type DNA
#residues 500-620 ##label SMI

```

```

#experimental_source strain K12
REFERENCE      A64720
#authors       Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
               Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
               Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
               Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
               Y.
#journal       Science (1997) 277:1453-1462
#title         The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession     G64725
#status        preliminary; nucleic acid sequence not shown;
               translation not shown
#molecule_type DNA
#residues      1-620 ##label BLAT
#cross-references GB:AE000115; GB:U000096; NID:g1786230; PID:g1786232;
               UMGp:b0047
#experimental_source strain K-12, substrain MG1655
GENETICS
#gene          kefc
#map_position  1 min
CLASSIFICATION #superfamily glutathione-regulated potassium efflux system
               protein kefc
KEYWORDS       potassium transport; transmembrane protein
FEATURE        48-70      #domain transmembrane #status predicted #label TM1\
               77-96      #domain transmembrane #status predicted #label TM2\
               103-124    #domain transmembrane #status predicted #label TM4\
               130-144    #domain transmembrane #status predicted #label TM5\
               149-170    #domain transmembrane #status predicted #label TM6\
SUMMARY        #length 620 #molecular-weight 67795 #checksum 3086
               9.4%; Score 92; DB 1; Length 620;
Query Match    Best Local Similarity 34.9%; Pred. No. 3.22e+00;
Matches 30; Conservative 19; Mismatches 27; Indels 10; Gaps 8;
Db 181 FALSALKVAGLVVLLGRVYTRPALRFVARSGLREVSVALFLV--FGFG-LLL-EE 236
QY 36 FTVS-LRDOAHAPVTLRASFADR-TLANISRWGKKLISAPFLFPQRF-FVLLLFQE 92
               ::::: ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
               ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
Db 237 VGLSMVANGAFLAGVLLASSEYRHALE 262
QY 93 GGLKLALNG--QG-LGATSMNQALE 115
               ::||: ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
               ::||: ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
Search completed: Tue Dec 15 15:54:53 1998
Job time : 23 secs.

```

W P E R E H (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search; using Smith-Waterman algorithm

Run on: Tue Dec 15 15:55:45 1998; MasPar time 10.89 seconds
608.412 Million cell updates/sec
Tabular output not generated.

Title: >US-09-109-864-2
Description: (1-133) from US09109864.pap
Perfect Score: 980

Sequence: 1 MSPRLVPCSHALPQGLSPG.....LEQLRELISGSVOLYCVHS 133

Scoring table: PAM 150
Gap 11

Searched: 165420 seqs, 49795644 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: spiremb16

1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 42.859; Variance 78.955; scale 0.543

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	146	14.9	354	11	36 KD BETA-GALACTOSIDE	6.48e-09
2	141	14.4	318	4	GALECTIN-8.	4.56e-08
3	113	11.5	136	11	GALECTIN-7.	1.46e-03
4	112	11.4	285	5	W09H1.68.	2.07e-03
5	111	11.3	262	13	GALECTIN-3.	2.93e-03
6	111	11.3	301	11	GALECTIN-6.	2.93e-03
7	110	11.2	283	5	GALECTIN.	4.14e-02
8	104	10.6	1262	5	F52H3.7.	3.18e-02
9	102	10.4	280	5	BETA-GALACTOSIDE-BINDI	6.19e-02
10	101	10.3	132	11	GALECTIN-7.	8.61e-02
11	100	10.2	195	14	BETA-A PROTEIN.	1.19e-01
12	98	10.0	573	10	TERPENE CYCLASE LIKE P	2.28e-01
13	98	10.0	603	11	COMPLEMENT COMPONENT F	2.28e-01
14	97	9.9	1309	10	REVERSE TRANSCRIPTASE.	3.15e-01
15	96	9.8	295	2	PUTATIVE INNER MEMBRAN	4.33e-01
16	95	9.7	703	6	ABC-TRANSPORTER.	5.95e-01
17	95	9.7	1320	5	SODIUM CHANNEL PROTEIN	5.95e-01
18	94	9.6	278	5	GALECTIN.	8.15e-01
19	92	9.4	166	2	HYPOTHETICAL 19.2 KD P	1.52e+00
20	92	9.4	278	5	GALECTIN.	1.52e+00

21	92	9.4	703	6	Q28435	ABC-TRANSPORTER.	1.52e+00
22	92	9.4	703	6	Q28437	ABC-TRANSPORTER.	1.52e+00
23	92	9.4	703	6	Q28436	ABC-TRANSPORTER.	1.52e+00
24	92	9.4	2124	2	Q53840	POLYKETIDE SYNTHASE (F	2.06e+00
25	91	9.3	260	5	O17270	R148.7 PROTEIN.	2.80e+00
26	90	9.2	324	2	O07208	HYPOTHETICAL 35.0 KD P	2.80e+00
27	90	9.2	928	14	P88954	ORF 63.	2.80e+00
28	89	9.1	64	5	Q29369	30 KDA S-TYPE LECTIN (3.79e+00
29	89	9.1	399	2	Q56257	HEMOLYSIN.	3.79e+00
30	89	9.1	862	10	Q43800	LIPOXYGENASE (EC 1.13.	5.11e+00
31	88	9.0	137	2	Q56973	KIM5.	5.11e+00
32	88	9.0	291	10	P93740	HYPOTHETICAL 32.4 KD P	5.11e+00
33	88	9.0	295	2	Q05769	HYPOTHETICAL 32.3 KD P	5.11e+00
34	87	8.9	198	14	O89433	COAT PROTEIN.	6.88e+00
35	87	8.9	198	14	O65721	COAT PROTEIN.	6.88e+00
36	87	8.9	346	1	O27843	CONSERVED PROTEIN.	6.88e+00
37	87	8.9	424	1	O26646	CATIONIC AMINO ACID TR	6.88e+00
38	87	8.9	445	1	O59214	445AA LONG HYPOTHETICA	6.88e+00
39	87	8.9	1670	5	Q23901	HISTIDINE KINASE.	6.88e+00
40	87	8.9	1711	2	P96311	ENDOGALACTINASE A (EC 3.	6.88e+00
41	86	8.8	309	2	O47455	PLASMID PRJ1004 DNA.	9.24e+00
42	86	8.8	353	8	O47472	CYTCHROME OXIDASE SUB	9.24e+00
43	86	8.8	403	1	O58825	403AA LONG HYPOTHETICA	9.24e+00
44	86	8.8	642	10	O64397	RECEPTOR-LIKE PROTEIN	9.24e+00
45	86	8.8	682	2	O67400	HYPOTHETICAL 77.8 KD P	9.24e+00

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	354 AA.
ID	O35866;			
AC	O35866;			
DT	01-JAN-1998 (TREMBREL. 05, CREATED)			
DT	01-JAN-1998 (TREMBREL. 05, LAST SEQUENCE UPDATE)			
DT	01-AUG-1998 (TREMBREL. 07, LAST ANNOTATION UPDATE)			
DE	36 KD BETA-GALACTOSIDE BINDING LECTIN.			
OS	RATTUS NORVEGICUS (RAT).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;			
OC	EUTHERIA; RODENTIA.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-SPRAGUE-DAWLEY; TISSUE-SMALL INTESTINE;			
RA	WADA J., OTA K., KUMAR A., KANWAR Y.S.;			
RL	SUBMITTED (SEP-1996) TO EMBL/GENBANK/DBJ DATA BANKS.			
DR	EMBL; U72741; G2351553;			
DR	PROSITE; P500309; LECTIN GALACTOSIDE; 2.			
DR	PFAM; PF00337; Gal-bind_lectin.			
KW	LECTIN.			
SQ	SEQUENCE 354 AA; 39946 MW; C7E7FEA CRC32;			

Query Match	14.9%;	Score 146;	DB 11;	Length 354;
Best Local Similarity	23.4%;	Pred. No. 6.48e-09;		
Matches	32;	Conservative 42;	Mismatches 57;	Indels 6; Gaps 4;
DB	219	TPAYTIPFTSPNGFPGSKSINISGVLPDAKRFINLR-C-GGDIAFHLPNRFNKKVVV	277	
QY	2	SPRLVPCSHALPQGLSPGVIIYRGVLQEPKHTVSLDQAHAPVTLRASFADETLA	61	
DB	278	RNTQINNSNGPERSLPGRMPNFGOSFVWILCEGHCFAVDGQHICFYHRLKNLPD	337	
QY	62	W---ISR-WG-QKKLISAPFLFYQRFVFLVLFQSGGLKALNGGIGATSMNQALEQ	116	
DB	338	INTLEVAGDIQLTHVQT	354	
QY	117	LRELISGSVOLYCVHS	133	
RESULT	2	PRELIMINARY;	PRT;	318 AA.
ID	O15215			
AC	O15215;			
DT	01-JAN-1998 (TREMBREL. 05, CREATED)			
DT	01-JAN-1998 (TREMBREL. 05, LAST SEQUENCE UPDATE)			
DT	01-JUN-1998 (TREMBREL. 06, LAST ANNOTATION UPDATE)			

```

DE GALECTIN-8.
GN GAL-8.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN HIPPOCAMPUS;
RA HADARI Y.R., EISENSTEIN M., ZAKUT R., ZICK Y.;
RL TRENDS GLYCOSCI. GLYCOTECHNOL. 9:103-112(1997).
DR EMBL; X91790; E199407; -.
DR PROSITE; PS00309; LECTIN_GALACTOSIDE; 1.
DR PFAM; PF00337; Gal-bind_lectin.
SQ SEQUENCE 318 AA; 35864 MW; 8FD5554C CRC32;

Query Match 14.4%; Score 141; DB 4; Length 318;
Best Local Similarity 24.6%; Pred. No. 4.56e-08;
Matches 34; Conservative 41; Mismatches 56; Indels 7; Gaps 5;

Db 181 TPOLRLPFAARLWTPMGPTVVVKGVEVNAKSNFVNDLLAGSKDIALHNLPRNLTKAF 240
QY 2 SPRLEVPCHALPOGLSPGVIIIRGLVQEPKHFVSL-RDQAAPVTLIRASFADRTL 60
Db 241 VNSFLOESWGSEERNITSPFSPGMYFEMITYCDYREKVVAVGVHSLEYKHFKE-LS 299
QY 61 AWIS--R--WGOKKLISAPFLFPQRFVFLVLLFQEGGLKALNG-OGLGATSMNQOALE 115
Db 300 SIDLTEIGHDIHLVRS 317
QY 116 QLRRLISGSVOLYCVHS 133

RESULT 3
ID O54974 PRELIMINARY; PRT; 136 AA.
AC O54974;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE GALECTIN-7.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C59 BLACK/6;
RX MEDLINE; 95246905.
RA MAGHALDO T., BERNERD F., DARMOH M.;
RL DEV. BIOL. 168:259-271(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C59 BLACK/6;
RA MAGNALDO T., FOWLES D., DARMOH M.;
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF038562; G2708654; -.
DR PROSITE; PS00309; LECTIN_GALACTOSIDE; 1.
SQ SEQUENCE 136 AA; 15173 MW; 61A0164B CRC32;

Query Match 11.5%; Score 113; DB 11; Length 136;
Best Local Similarity 27.7%; Pred. No. 1.46e-03;
Matches 36; Conservative 37; Mismatches 44; Indels 13; Gaps 9;

Db 9 SLPGQVRVTVNRIRGLVDPQAGRPVNLGCEGADAAALHFNPRDLTSEVFNTRKQG 68
QY 12 ALPOGLSPGVIIIRGLVQEPKHFVSL---RDQAAPVTLIRASFADRTLAW-I---S 64
Db 69 KNGREGRGIPQR-GOPF-EVLLIATEEGFK-AVVGDD-EYLHFNHRLPPARVRLVEY 124
QY 65 RRG-OKKLISAPFLFPQRFVFLVLLFQEGGLKALNG-OGLGATSMNQO-ALQLRLRI 122
Db 125 GGDVQLHSLN 134
QY 123 SGVOLYCVH 132

RESULT 4
ID O45904 PRELIMINARY; PRT; 285 AA.
AC O45904;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE W09H1.68.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNITEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RA SNEY R.;
RL SUBMITTED (JAN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAYTON M.,
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FUTON L., GARDNER A., GREEN P.,
RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA KIRSTEN J., LAISTER H., LATREILLE P., LIGHTNING J., LLOYD C.,
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,
RA SONNHAMMER E., STADEH R., SULSTON J., THETTERY-MIEG J., THOMAS K.,
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA WILKINSON-SPROAT J., WOHLDMAN P.;
RL NATURE 368:32-38(1994).
DR EMBL; Z82081; E1247370; -.
DR PROSITE; PS00309; LECTIN_GALACTOSIDE; 2.
SQ SEQUENCE 285 AA; 32683 MW; 920E5744 CRC32;

Query Match 11.4%; Score 112; DB 5; Length 285;
Best Local Similarity 35.9%; Pred. No. 2.07e-03;
Matches 14; Conservative 12; Mismatches 13; Indels 0; Gaps 0;

Db 17 VPYRSVLOERFEPQTLIVKSGSIDESQRTINLHSTKA 55
QY 7 VPCSHALPOGLSPGVIIIRGLVQEPKHFVSLRDQA 45

RESULT 5
ID Q90713 PRELIMINARY; PRT; 262 AA.
AC Q90713;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE GALECTIN-3.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGATHAE;
OC GALLIFORMES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CARTILAGE;
RA NURMINSKAYA M.V., LINSSENMEYER T.F.;
RL DEV. DYN. 0:0-0(0).
DR EMBL; U50339; G1389600; -.
DR PFAM; PF00337; Gal-bind_lectin.
KW LECTIN.
SQ SEQUENCE 262 AA; 28155 MW; FD6F1E17 CRC32;

Query Match 11.3%; Score 111; DB 13; Length 262;
Best Local Similarity 25.0%; Pred. No. 2.93e-03;
Matches 34; Conservative 30; Mismatches 66; Indels 6; Gaps 5;

Db 122 AAPLKVYDPLPAGLMPRLITITGVNSNPNRFSLDKRGODIAPHFPRFREDHKRV 181
QY 2 SPRLEVPCHALPOGLSPGVIIIRGLVQEPKHFVSL-RDQA-A-HAPVTLIRASFAD 58
Db 152 IVNSMFQNWNGREERTAPRFPFGTFFKLVQECEDHFKVAVNDALHLOFNPFEKLN 241
QY 59 TL-AWI--SRWGOKKLISAPFLFPQRFVFLVLLFQEGGLKALNG-OGLGATSMNQOALE 115
Db 242 GITKICLAGDITLTSV 257

```

Query Match	Score 110;	DB 5;	Length 283;
Best Local Similarity 35.9%;	Pred. No. 4.14e-03;		
Matches 14;	Conservative 12;	Mismatches 13;	Indels 0;
Gaps 0;			

```

RN RP SEQUENCE FROM N.A.
RA BEVAN M., TERRY N., ARDILES W., BUYSSHAERT C., DASSEVILLE R.,
DE CLERCK R., DE KEYSER A., NEYT P., ROUZE P., VAN DEN DAEL H.,
RA VILLAROEEL R., GIELEN J., VAN MONTAGU M., BANCROFT I., MEWES H.W.,
RA MAYER K., SCHUELLER C.,
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
[2]
RN RP SEQUENCE FROM N.A.
RA EU ARABIDOPSIS SEQUENCING PROJECT:
RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AL0283969; -
SQ SEQUENCE 573 AA; 65617 MW; 21B69D3A CRC32;

Query Match 10.0%; Score 98; DB 10; Length 573;
Best Local Similarity 28.2%; Pred. No. 2.28e-01;
Matches 22; Conservative 20; Mismatches 31; Indels 5; Gaps 5

Db 213 KDFILEALGFTSMLESFSLAASGTCPPHLSVRIRNALGLSQHNWMLVPVEFIPFYEQE 272
Oy : : : : : : : : : : : : : : : : : : : : : : : : : : : :
26 RGLVLOEPKHFVS-LRDAQAHAPVTLRASFAADR-TLAWISRWGOKKLISAPFL-FYPQR 82
Oy : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 273 IEHDEMLLKFAKLSFKLG 290
Oy : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Oy 83 F-FEVILL-FQEGGLKLA 98

RESULT 13
ID Q61129 PRELIMINARY; PRT; 603 AA.
AC Q61129;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DI 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE DE COMPLEMENT COMPONENT FACTOR I (COMPLEMENT FACTOR I).
GN CFI.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA: VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RX MEDLINE; 96175003.
RA MANTA J.O., WONG M.J., KOZAK C.A., KUNNATH-MUGLIA L.M., GOLDBERGER G.;
RL MOL. IMMUNOL. 33:101-112(1996).
DR EMBL: U47810; G1322034; -
DR MGD; MGI:105937; CFI.
DR PGD; PS01209; LDLRA_1; 1.
DR PFAM: PF00057; ldl_recept.a.
DR PFAM: PF00089; trypsin.
DR PFAM: PF00530; SRCR.
KW GLYCOPROTEIN.
SQ SEQUENCE 503 AA; 67289 MW; F3502F5F CRC32;

Query Match 10.0%; Score 98; DB 11; Length 603;
Best Local Similarity 47.6%; Pred. No. 2.28e-01;
Matches 10; Conservative 5; Mismatches 6; Indels 0; Gaps 0

Db 505 RWGEVDLIGNCSOFPDRIYE 525
Oy : : : : : : : : : : : : : : : : : : : : : : : : : : : :
65 RWGOKKLISAPFLFYPQRFEE 85

RESULT 14
ID O64428 PRELIMINARY; PRT; 1309 AA.
AC O64428;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DI 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE REVERSE TRANSCRIPTASE.
OS CHLORELLA FLAVIGIS.
OC EUKARYOTA; PLANTAE; PHYCOPHYTA; CHLOROPHYTA (GREEN ALGAE);
OC CHLOROPHYCEAE; CHLOROCCOCALLES; OOCYSTACEAE.
RN [1]

```

Search completed: Tue Dec 15 15:56:10 1998
Job time : 25 secs.